

LONDON
SCHOOL of
HYGIENE
& TROPICAL
MEDICINE



LSHTM Research Online

Pickering, H; (2017) Identification of Chlamydia trachomatis Immune Targets through Immunological and Population-Genomic Screens and Elucidation of Potential Roles in Bacterial Pathogenesis. PhD thesis, London School of Hygiene & Tropical Medicine. DOI: <https://doi.org/10.17037/PUBS.03928322>

Downloaded from: <https://researchonline.lshtm.ac.uk/id/eprint/3928322/>

DOI: <https://doi.org/10.17037/PUBS.03928322>

Usage Guidelines:

Please refer to usage guidelines at <https://researchonline.lshtm.ac.uk/policies.html> or alternatively contact researchonline@lshtm.ac.uk.

Available under license. To note, 3rd party material is not necessarily covered under this license: <http://creativecommons.org/licenses/by-nc-nd/3.0/>

<https://researchonline.lshtm.ac.uk>

LONDON
SCHOOL of
HYGIENE
& TROPICAL
MEDICINE



**Identification of *Chlamydia trachomatis* Immune Targets through
Immunological and Population-Genomic Screens and Elucidation of
Potential Roles in Bacterial Pathogenesis**

HARRY CHARLES PICKERING

**Thesis submitted in accordance with the requirements for the degree
of**

**Doctor of Philosophy of the
University of London**

SEPTEMBER 2016

Department of Clinical Research

Faculty of Infectious and Tropical Diseases

LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE

Funded by The Bloomsbury Colleges

Research group affiliation(s): Trachoma group

Declaration

I, Harry Pickering, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed:

A handwritten signature in black ink, appearing to read 'H. Pickering', written in a cursive style.

Date: 26/09/2016

Harry Pickering

Abstract

Ocular infection by *Chlamydia trachomatis* (Ct) results in trachoma, the leading infectious cause of blindness. Infection clears naturally, but repeated exposure in endemic areas and resulting inflammation promote tissue damage leading eventually to blinding sequelae. Antibiotic treatment as part of community-based intervention reduces prevalence of infection and disease but rarely eliminates the problem completely and progression to scarring and blindness does still occur. Sixty years of vaccine trials have produced variable results therefore new candidate antigens and better understanding of the underlying causes of infection and disease are required.

Serum samples from trachoma-endemic communities in The Gambia were tested against the arrayed Ct proteome to identify antibody responses associated with protection from infection and from scarring disease. More focussed global antibody profiles were associated with partial immunity to infection. Several antibody targets were identified as individually associated with infection and disease outcome. Clinical Ct isolates collected from Guinea-Bissau were screened for evidence of natural selection to identify further immune targets and to validate those discovered through serological techniques. Evidence of positive selection was found for known Ct virulence factors, there was little evidence of balancing selection. Antibody targets associated with susceptibility to infection and scarring had evidence of purifying selection. One of the Ct antigens, CT442, identified as being an immune target and under natural selection was characterised further using cell-culture models. It was localised to the inclusion membrane through immunofluorescence microscopy, the primary point of contact with the host, and potentially interacted with pathways involved in intracellular vesicular trafficking based on interacting proteins identified through mass spectrometry.

Ct infection is shown to stimulate a broad, polyclonal antibody response, individuals with more focussed responses are better protected from persistent infection and scarring progression. Purifying selection in antibody targets which associate with poor resolution of infection suggests two possible hypotheses for Ct evasion of immune responses. The decoy hypothesis, in which Ct actively promotes immune responses against irrelevant, decoy antigens to divert antibody responses away from protective antigens, and the blocking hypothesis, in which antibodies against certain Ct surface antigens block the binding of neutralising antibodies. Evidence of selection in CT442 show it is important but unlikely essential for Ct survival, the functions that are driving this evolution require further study.

Acknowledgements

This project owes a great debt to my supervisors Martin Holland and Richard Hayward. From initial conception to continued guidance during these three years, I'd like to thank them both for the opportunity to do this work and for their support and probably patience as well.

Everyone in the office; Bert, Christine, Jo, Sandra, Steph, Tamsyn and any others who've passed through. Thanks for plenty of advice and making my time here undoubtedly easier and a lot of fun. I'd also like to thank the extended Trachoma group, particularly Chrissy Roberts who has answered questions on just about every subject covered in this thesis and many not.

I'd also like to mention Charlotte, Ellie, Maud and Ollie at Birkbeck. Thanks for all your help in the lab and on the microscope, it was great fun working with all of you.

I would also like to thank the many members of the Trachoma group and the field teams in Guinea-Bissau, Tanzania and The Gambia who performed the studies and collected the samples used in this project. This thesis encompasses studies dating back 21 years and the results presented owe a massive amount to the countless numbers of people involved in obtaining and keeping a hold of these samples, and the many participants who agreed to be involved.

Finally, I have to and am delighted to thank my mum and dad, for everything between Coventry and here.

Table of Contents

ABSTRACT.....	3
ACKNOWLEDGEMENTS.....	4
LIST OF FIGURES.....	11
LIST OF TABLES.....	15
ABBREVIATIONS.....	17
1. INTRODUCTION	18
1.1. TRACHOMA	19
1.1.1. Epidemiology	19
<i>Image courtesy of the International Coalition for Trachoma control</i> <i>(http://www.trachomacoalition.org/)</i>	21
1.1.2. Control and elimination	22
1.1.3. Risk factors for trachoma.....	22
1.2. CHLAMYDIA TRACHOMATIS	23
1.2.1. History and evolution within and between species.....	23
1.2.2. Host cell entry and early host interactions	26
1.2.3. Chlamydial inclusion and its host interactions.....	30
1.2.4. Exit and the potential for persistence	32
1.2.5. Chlamydia trachomatis population genetics	32
1.3. IMMUNOLOGY OF TRACHOMA	34
1.3.1. Initial responses to Ct infection.....	34
1.3.2. Immunopathology of recurrent Ct infection and persistent inflammation	35
1.4. IMMUNE TARGETS IN CT INFECTION	36
1.4.1. Early studies of immunodominant antigens	36
1.4.2. The history and importance of protein-array screens for serological profiling.....	36
1.4.3. Protein-array screens of serological and cellular responses from persons with Ct infection..	37
1.4.4. Serological screen of Gambian trichiasis patients and healthy controls.....	38
1.5. VACCINE TRIALS AND CURRENT STATUS.....	40
1.5.1. MOMP as the primary vaccine candidate	41
1.5.2. New targets and delivery systems	41
1.5.3. Current status	42
2. HYPOTHESES AND RESEARCH QUESTIONS.....	43
2.1. HUMORAL IMMUNITY PLAYS A SIGNIFICANT ROLE IN PROTECTION FROM HUMAN OCULAR CHLAMYDIA TRACHOMATIS INFECTION AND TRACHOMATOUS DISEASE.....	44
2.1.1. Do serological anti-Chlamydia trachomatis responses correlate with resolution of infection and protection from reinfection?.....	44

2.1.2. Do serological responses associate with the blinding sequelae of infection in trachoma?	44
2.1.3. Do antibodies limit progression of scarring in endemic communities?	44
2.2. TARGETS OF HUMORAL IMMUNITY IN TRACHOMA ARE UNDER NATURAL SELECTION	44
2.2.1. Can we identify new antigenic targets of Ct immunity in trachoma using protein microarrays?	44
2.2.2. Are the identified antibody targets under immune-driven selection?	44
2.2.3. Are signatures of selection detectable within a population of ocular Ct strains?	44
2.3. CT442 IS A HOST IMMUNE TARGET IN TRACHOMA THAT IS INVOLVED IN BACTERIAL PATHOGENESIS	44
2.3.1. What is the intracellular localisation of CT442 and does this facilitate intracellular development and survival of Ct through specific host interactions?	44
2.3.2. What host and/or chlamydial proteins and pathways does CT442 engage?	44
3. METHODS	45
3.1. PROTEIN EXPRESSION AND PURIFICATION	46
3.1.1. Agarose and SDS-polyacrylamide gel electrophoresis	46
PROTEIN SAMPLES DILUTED 1:1 IN LAEMMLI BUFFER AND INCUBATED AT 95°C FOR 10 MINUTES, WERE VISUALISED BY SODIUM DODECYL SULPHATE (SDS)-POLYACRYLAMIDE GEL ELECTROPHORESIS USING EITHER NUSEP PRE-CAST PROTEIN GELS (HOMEBUSH, NSW) OR IN-HOUSE GELS FOR 1.5 TO 3 HOURS USING 120 TO 140 VOLTS. BANDS WERE VISUALISED USING COOMASSIE BLUE DYE OR SILVER STAIN, WHEN THE PROTEIN CONCENTRATION WAS EXPECTED TO BE < 0.2 MG.ML. BAND SIZES WERE ESTIMATED USING EITHER PRECISION PLUS PROTEIN DUAL COLOUR (BIO-RAD) OR AMERSHAM ECL FULL-RANGE RAINBOW (GE HEALTHCARE LIFE SCIENCES) MOLECULAR WEIGHT MARKER.	46
3.1.2. Immunoblotting	46
3.1.3. Provided Ct constructs	47
3.1.4. Transformation and miniprep of provided constructs	48
3.1.5. Design and production of constructs	48
3.1.6. Optimisation of expression and purification conditions	50
3.1.7. Peptide selection and production	52
3.2. SEROLOGICAL SCREENING OF PROTEOME ARRAYS	52
3.2.1. Glutathione S-transferase fusion protein micro-titre plate array	52
3.2.2. Microarray chips	54
3.2.3. Normalisation and filtering	54
3.2.4. Determining positivity	55
3.2.5. Choice of outcome variable and covariates	57
3.2.6. Diversity metrics	58
3.3. IN SILICO ANALYSES	61
3.3.1. Expression, localisation and structure	61
3.3.2. B and T-cell epitopes predictions	62
3.4. STUDY DETAILS OF TESTED SERUM SAMPLES	62
3.4.1. HLA polymorphism and scarring trachoma	62
3.4.2. NK-cells and trachoma	63

3.4.3. Longitudinal cohort study of Ct infection and active trachoma.....	63
3.4.4. Scarring case-control study.....	64
3.4.5. NK-cells and scarring trachoma.....	65
3.4.6. Longitudinal cohort in Tanzanian children investigating scarring progression (2012-2016)..	65
3.4.7. NK-cells and HCMV, cross-sectional survey (The Gambia 2011).....	67
3.4.8. The Solomon Islands Trachoma prevalence survey 2014.....	67
3.5. HUMORAL IMMUNOLOGY.....	68
3.5.1. Optimisation and validation of the in-house ELISA.....	68
3.5.2. Serum/plasma ELISA.....	68
3.5.3. Dried blood-spot ELISA.....	69
3.5.4. Biotinylated peptide ELISA.....	69
3.5.5. Between-plate normalisation.....	70
3.5.6. Comparison of OD values.....	70
3.6. POPULATION GENOMICS.....	70
3.6.1. Isolate collection and whole-genome sequencing.....	70
3.6.2. Alignment, assembly and filtering by individual genes.....	72
3.6.3. Allele frequency-based signatures of selection.....	72
3.6.4. Haplotype-based signatures of selection.....	75
3.7. CELL BIOLOGY USING CELL-CULTURE MODELS.....	78
3.7.1. Mammalian cell-culture and Ct infection.....	78
3.7.2. GFP-construct transfection.....	78
3.7.3. Fixation and antibody staining for IF microscopy.....	78
3.7.4. Co-Immunoprecipitation of CT442-GFP from HeLa cells.....	79
3.7.5. Liquid chromatography-mass spectrometry.....	80
3.7.6. Filtering of mass spectrometry hits.....	80
3.7.7. Pathway-enrichment analysis.....	81
4. ASSOCIATION BETWEEN ANTIBODY RESPONSE AND LONGITUDINAL OCULAR INFECTION	83
4.1. INTRODUCTION.....	84
4.1.1. Natural history of trachoma.....	84
4.1.2. A balance of protective and pathogenic immune responses.....	84
4.1.3. Proteome-wide antibody profiling.....	86
4.1.4. Study design and arrayed sera.....	86
4.2. RESULTS.....	87
4.2.1. Protection from Ocular Ct infection.....	87
4.2.2. Array normalisation and filtering.....	90
4.2.3. More focussed global antibody responses are associated with protection.....	93
4.2.4. Individual antibody responses are associated with a lack of protection.....	96
4.2.5. Combinatorial antibody responses identified by multivariate regression do not reliably predict susceptibility to infection.....	105

4.2.6. Individuals with a more focussed antibody profile are associated with protection.....	108
4.2.7. Selective ELISA testing of differentially recognised antigens.....	110
4.2.8. ELISA validation showed variable association with a lack of protection	117
4.3. DISCUSSION.....	131
4.3.1. Longitudinal evidence of Ct infection reflects differences in immunity	131
4.3.2. More focussed antibody responses protect from acquisition and long duration of infection	132
4.3.3. Higher responses to 42 antigens were associated with susceptibility to infection	133
4.3.4. Presentation of non-surface antigens to the humoral immune system.....	135
4.3.5. Limited identification of antigens associated with protection.....	136
4.4. CONCLUSIONS AND FUTURE WORK	138
5. GLOBAL PROFILING OF CHLAMYDIA TRACHOMATIS-SPECIFIC ANTIBODY RESPONSES IN TRACHOMATOUS TRICHIASIS.....	140
5.1. INTRODUCTION	141
5.1.1. Progressive scarring in trachoma and associated risk factors.....	141
5.1.2. Dysregulation of inflammation and wound healing promote scarring.....	141
5.1.3. Association between scarring and antibody responses	142
5.1.4. Study design and initial analysis	143
5.2. RESULTS	145
5.2.1. Array Normalisation and filtering	145
5.2.2. Antibody responses were focussed on proteins expressed early and late during the developmental cycle and localised to interact with the host.....	148
5.2.3. Individual antibody responses are associated with cross-sectional evidence of conjunctival scarring	154
5.2.4. Combinatorial antibody responses identified by multivariate regression did not improve predictions of conjunctival scarring	158
5.2.5. Selective expression identified antigens and independent repeat ELISA testing of differentially recognised antigens.....	161
5.2.6. ELISA validation of antibody responses associated with scarring.....	168
5.2.7. Progression of scarring in a longitudinal cohort	175
5.3. DISCUSSION.....	178
5.3.1. Adults have more focussed antibody responses	179
5.3.2. Few antigens were differentially recognised between individuals with and without scarring trachoma	180
5.3.3. CT442 and CT706 antibody responses are associated with differential scarring progression	182
5.4. CONCLUSIONS AND FUTURE WORK	183

6. IDENTIFICATION OF <i>CHLAMYDIA TRACHOMATIS</i> GENES UNDER SELECTION USING WHOLE GENOME SEQUENCING AND CORRELATION WITH SEROLOGICAL IMMUNOGENICITY	185
6.1. NATURAL SELECTION, PATHOGEN ADAPTATION AND ANTIGENIC VARIATION	186
6.1.1. <i>Natural selection as an evolutionary process</i>	<i>186</i>
6.1.2. <i>Antigenic diversity as natural selection</i>	<i>187</i>
6.1.3. <i>Markers of selection in human pathogens</i>	<i>187</i>
6.1.4. <i>Genes under selection in Chlamydia trachomatis</i>	<i>189</i>
6.2. RESULTS	191
6.2.1. <i>Prediction of immunogenicity</i>	<i>191</i>
6.2.2. <i>Genome-wide evidence of purifying and positive selection, with minimal balancing selection</i>	<i>194</i>
6.2.3. <i>Integrated haplotype scores identify three genomic regions under positive selection</i>	<i>208</i>
6.2.4. <i>Variable evidence of selection acting on target genes</i>	<i>214</i>
6.2.5. <i>Evidence of selection in predicted epitopes</i>	<i>219</i>
6.3. DISCUSSION	224
6.3.1. <i>Genome-wide evidence of purifying and positive selection</i>	<i>224</i>
6.3.2. <i>Secreted proteins, outer and inclusion membrane proteins are under strong forces of selection</i>	<i>225</i>
6.3.3. <i>Limited evidence of balancing selection and Ct immune evasion</i>	<i>227</i>
6.3.4. <i>Evidence of selection in susceptibility and scarring associated antigens identifies candidate immune targets</i>	<i>228</i>
6.4. CONCLUSIONS AND FUTURE WORK	230
7. FUNCTIONAL STUDIES OF CT442 A CANDIDATE ANTIGEN ASSOCIATED WITH PROTECTION FROM SCARRING TRACHOMA.....	231
7.1. INTRODUCTION	232
7.1.1. <i>Functional importance of inclusion membrane proteins</i>	<i>232</i>
7.1.2. <i>Limited previous characterisation of CT442</i>	<i>232</i>
7.1.3. <i>CT442 is a cellular and humoral immune target under natural selection</i>	<i>233</i>
7.1.4. <i>Polymorphism and selection in CT442</i>	<i>234</i>
7.2. RESULTS	236
7.2.1. <i>CT442 sequence homology and predicted structure</i>	<i>236</i>
7.2.2. <i>CT442-His construct design and expression</i>	<i>239</i>
7.2.3. <i>CT442-GFP localised to the endoplasmic reticulum</i>	<i>246</i>
7.2.4. <i>CT442 localised in the inclusion membrane from mid-cycle onwards</i>	<i>251</i>
7.2.5. <i>Mass-spectrometry of CT442-GFP Co-IP identified potential interactors</i>	<i>256</i>
7.2.6. <i>Pathway enrichment and network analyses highlight intracellular vesicular trafficking</i>	<i>265</i>
7.2.7. <i>Chlamydial targets link with human pathway enrichment of vesicular trafficking</i>	<i>274</i>
7.2.8. <i>Confirmation of Rab7 immunoprecipitated by CT442-GFP</i>	<i>274</i>

7.3. DISCUSSION.....	275
7.3.1. CT442 is polymorphic and it is not essential for Ct survival	276
7.3.2. CT442 localised to the inclusion membrane from 24 hours post infection	277
7.3.3. Co-immunoprecipitation of CT442-GFP from Ct infected cells identifies intracellular trafficking and Rab proteins	277
7.3.4. Rab proteins and Chlamydia trachomatis.....	279
7.3.5. Lipid droplets and Chlamydia trachomatis	280
7.4. CONCLUSIONS AND FUTURE WORK	281
8. DISCUSSION	282
8.1. NATURAL SELECTION WITHIN A POPULATION OF OCULAR CT ISOLATES	283
8.1.1. Selection validates Ct host-interactors	283
8.1.2. Selection identifies genes coding for immunogenic proteins.....	285
8.2. ANTIBODY RESPONSES AND SUSCEPTIBILITY TO CT INFECTION.....	289
8.2.1. The blocking hypothesis.....	290
8.2.2. The decoy hypothesis.....	293
8.2.3. Limitations of the natural immune response to ocular Ct infection	297
8.3. CT442, RAB PROTEINS AND INTRACELLULAR TRAFFICKING	298
8.3.1. CT442 is evolving under non-neutral selection	299
8.3.2. Potential interactions of CT442	300
8.4. IMPLICATIONS	301
8.4.1. Implications for a Chlamydia trachomatis vaccine	301
8.4.2. Implications for Ct genomics.....	303
8.5. LIMITATIONS.....	304
8.5.1. Microarray limitations	304
8.5.2. Influence of urogenital chlamydial infection	304
8.5.3. Interplay of humoral and cellular immunity	305
8.6. FUTURE WORK	307
8.6.1. Serological validation.....	307
8.6.2. Independent populations for Ct genomics	308
8.6.3. Understanding the role of CT442 in the Ct developmental cycle.....	308
8.7. SUMMARY.....	309
APPENDIX	335
BUFFERS	335

List of Figures

FIGURE 1.1. WHO SIMPLIFIED TRACHOMA GRADING SYSTEM.	20
FIGURE 1.2. GLOBAL MAP OF ACTIVE TRACHOMA (2015).	21
FIGURE 1.3. SAFE STRATEGY FOR TRACHOMA CONTROL.	22
FIGURE 1.4. MAXIMUM LIKELIHOOD RECONSTRUCTION OF THE CT WHOLE-GENOME PHYLOGENY.	25
FIGURE 1.5. CT DEVELOPMENTAL CYCLE.	27
FIGURE 3.1. VISUALISATION OF THEORETICAL ANTIBODY RESPONSES IN TWO SAMPLES, DIFFERENT COLOURS REPRESENT ABUNDANCE OF ANTIBODY RESPONSES TO DIFFERENT ANTIGENS.	61
FIGURE 3.2. GENETIC DISTANCE-BASED IMPUTATION OF MISSING CALLS.	77
FIGURE 4.1. VISUALISATION OF LONGITUDINAL EVIDENCE OF OCULAR CT INFECTION IN 90 INDIVIDUALS DURING THE SIX- MONTH STUDY.	89
FIGURE 4.2. TRANSFORMATION AND NORMALISATION OF THE RAW ARRAY DATA.	91
FIGURE 4.3. AVERAGE SILHOUETTE WIDTHS FOR CLUSTERING METHOD TRIALLED FOR ALL 441 ANTIGENS.	92
FIGURE 4.4. INCREASED BREADTH OF RESPONSES IN SUSCEPTIBLE INDIVIDUALS.	93
FIGURE 4.5. INCREASED DIVERSITY AND LESS FOCUSED RESPONSES IN SUSCEPTIBLE INDIVIDUALS.	95
FIGURE 4.6. OVER-REPRESENTATION OF LATE AND VERY EARLY EXPRESSED PROTEINS IN ANTIGENS ASSOCIATED WITH SUSCEPTIBILITY.	99
FIGURE 4.7. OVER-REPRESENTATION OF PROTEINS EXTRACELLULAR, INNER MEMBRANE AND PERIPLASM IN ANTIGENS ASSOCIATED WITH SUSCEPTIBILITY.	100
FIGURE 4.8. ANTIBODY RESPONSES AVERAGED ACROSS THE TOP 32 DIFFERENTIALLY RECOGNISED ANTIGENS.	103
FIGURE 4.9. COMBINATORIAL ANTIBODY RESPONSES DID NOT ACCURATELY PREDICT SUSCEPTIBILITY.	104
FIGURE 4.10. RANKED VARIABLE IMPORTANCE FROM RANDOM FORESTS REGRESSION.	106
FIGURE 4.11. MULTIVARIATE REGRESSION DID NOT IDENTIFY ANTIGENS ABLE TO PREDICT SUSCEPTIBILITY.	107
FIGURE 4.12. DIFFERENTIALLY RECOGNISED ANTIGENS FROM THE COMPLETE SET, 'GLOBALLY HIGHER' INDIVIDUALS AND 'GLOBALLY LOWER' INDIVIDUALS HAD LIMITED OVERLAP.	109
FIGURE 4.13: COOMASSIE BLUE STAINED SDS-PAGE GEL OF CT089 AND CT875.	111
FIGURE 4.14. EXPRESSION TRIALS OF CT795-GST AND CT813-GST.	113
FIGURE 4.15: EXPRESSE 8/7, AGGREGATE MOSTLY 7, AGGREGATE AND INSOLUBLE 8	113
FIGURE 4.16. EXPRESSION AND PURIFICATION OF PGP3	116
FIGURE 4.17. OPTIMISATION OF AN IN-HOUSE ELISA PROTOCOL.	118
FIGURE 4.18. VALIDATION OF AN IN-HOUSE PGP3 ELISA.	120
FIGURE 4.19. CT089 CORRELATION BETWEEN ELISA AND ARRAY RESULTS.	122
FIGURE 4.20. CT875 CORRELATION BETWEEN ELISA AND ARRAY RESULTS.	123
FIGURE 4.21. INCG CORRELATION BETWEEN ELISA AND ARRAY RESULTS.	124
FIGURE 4.22. INCA CORRELATION BETWEEN ELISA AND ARRAY RESULTS.	125
FIGURE 4.23. PGP3 RESPONSES IN 90 ARRAYED SERA.	126
FIGURE 4.24. ELISA RESPONSES IN "BOTH DISCORDANT" SERA WERE LOWER FOR ALL ANTIGENS.	129
FIGURE 5.1. TRANSFORMATION AND NORMALISATION OF THE RAW ARRAY DATA.	146
FIGURE 5.2. AVERAGE SILHOUETTE WIDTHS FOR CLUSTERING METHOD TRIALLED FOR ALL 441 ANTIGENS.	147

FIGURE 5.3. OVER-REPRESENTATION OF LATE AND VERY EARLY EXPRESSED PROTEINS IN IMMUNOGENIC ANTIGENS.....	149
FIGURE 5.4. OVER-REPRESENTATION OF PROTEINS EXTRACELLULAR, OUTER MEMBRANE AND PERIPLASM IN IMMUNOGENIC ANTIGENS.....	150
FIGURE 5.5. NO DIFFERENCE IN BREADTH OF ANTIGENS RECOGNISED BETWEEN ADULTS WITH AND WITHOUT SCARRING. ...	151
FIGURE 5.6. NO SIGNIFICANT DIFFERENCES IN DIVERSITY OF ANTIBODY RESPONSES BETWEEN ADULTS WITH AND WITHOUT SCARRING.....	153
FIGURE 5.7. DIFFERENTIAL ANTIBODY RESPONSES BETWEEN ADULTS WITH AND WITHOUT SCARRING.	155
FIGURE 5.8. COMBINATORIAL ANTIBODY RESPONSES DID NOT ACCURATELY PREDICT PRESENCE OF SCARRING.	157
FIGURE 5.9. RANKED VARIABLE IMPORTANCE FROM RANDOM FORESTS REGRESSION.	159
FIGURE 5.10. MULTIVARIATE REGRESSION IDENTIFIED ANTIGENS WITH MODEST PREDICTIVE POWER.	160
FIGURE 5.11. EXPRESSION TRIALS OF CT442-GST.....	162
FIGURE 5.12. EXPRESSION TRIALS OF CT442N-GST AND CT442C-GST.	164
FIGURE 5.13. EXPRESSION TRIALS OF CT667-GST.....	166
FIGURE 5.14. EXPRESSION TRIALS OF CT706-GST.....	167
FIGURE 5.15. PGP3 CORRELATION BETWEEN ELISA AND ARRAY RESULTS.	169
FIGURE 5.16. CT442 CORRELATION BETWEEN ELISA AND MICROTITRE PLATE ARRAY RESULTS.....	170
FIGURE 5.17. CT667 CORRELATION BETWEEN ELISA AND ARRAY RESULTS.	171
FIGURE 5.18. CT706 CORRELATION BETWEEN ELISA AND ARRAY RESULTS.	172
FIGURE 5.19. ELISA RESPONSES AGAINST CT442, CT667 AND CT706 IN THE COMPLETE SET OF 116 SERA.....	173
FIGURE 5.20. CT442 ANTIBODY RESPONSES WERE SIGNIFICANTLY HIGHER IN CHILDREN WHOM CONJUNCTIVAL SCARRING HAD NOT PROGRESSED.	177
FIGURE 5.21. CT706 ANTIBODY RESPONSES WERE SIGNIFICANTLY HIGHER IN CHILDREN WHOM CONJUNCTIVAL SCARRING HAD PROGRESSED.....	178
FIGURE 6.1. HISTOGRAM OF THE GENOME-WIDE DISTRIBUTION OF TAJIMA'S D.	196
FIGURE 6.2. GENOME-WIDE DISTRIBUTIONS OF TAJIMA'S D, Fu AND Li'S D* AND F*.....	197
FIGURE 6.3. GENOME-WIDE EVIDENCE OF SELECTION BY TAJIMA'S D.	199
FIGURE 6.4. GENOME-WIDE DISTRIBUTION OF FAY AND WU'S H.	203
FIGURE 6.5. CORRELATION OF TAJIMA'S D AND FAY AND WU'S H.....	204
FIGURE 6.6. CORRELATION OF TAJIMA'S D AND FAY AND WU'S H FROM SLIDING-WINDOW ANALYSIS.	207
FIGURE 6.7. GENOME-WIDE EVIDENCE OF SNPs AND LOCI UNDER POSITIVE SELECTION BY IHS.....	210
FIGURE 6.8. OVERLAP OF REGIONS UNDER SELECTION BY TAJIMA'S D OR FAY AND WU'S H WITH IHS.	211
FIGURE 6.9. OVER-REPRESENTATION OF GENES EXPRESSED EARLY OR LATE IN THE DEVELOPMENTAL CYCLE AND LOCALISED TO THE OUTER MEMBRANE OR SECRETED IN GENES UNDER SELECTION.....	213
FIGURE 6.10. EVIDENCE OF SELECTION IN ANTIBODY TARGETS ASSOCIATED WITH PROTECTION FROM OR SUSCEPTIBILITY TO INFECTION.	215
FIGURE 6.11. EVIDENCE OF SELECTION IN ANTIBODY TARGETS ASSOCIATED THE PRESENCE OR LACK OF SCARRING IN ADULTS.	216
FIGURE 6.12. EVIDENCE OF SELECTION IN CT681 (<i>OMPA</i>).	220
FIGURE 6.13. EVIDENCE OF SELECTION IN CT868 (<i>DUB1</i>).	221

FIGURE 6.14. EVIDENCE OF SELECTION IN CT442 (CRPA).	222
FIGURE 6.15. EVIDENCE OF SELECTION IN CT228.	223
FIGURE 6.16. EVIDENCE OF SELECTION IN CT694.	223
FIGURE 7.1. SCHEMATIC OF THE PREDICTED STRUCTURE OF CT442.	236
FIGURE 7.2. AMINO ACID ALIGNMENT OF CT442 FROM SEROVAR-REPRESENTATIVE STRAINS.	238
FIGURE 7.3. EXPRESSION TRIALS OF CT442-HIS.	240
FIGURE 7.4. INDUCTION TRIALS OF CT442-HIS.	241
FIGURE 7.5. FURTHER INDUCTION TRIALS OF CT442-HIS.	242
FIGURE 7.6. SIZE EXCLUSION-BASED CHROMATOGRAPHY FOR CT442-HIS PURIFICATION.	244
FIGURE 7.7. CO-EXPRESSION OF CT442-HIS WITH MOLECULAR CHAPERONES.	245
FIGURE 7.8. LOCALISATION OF GFP AND CT442-GFP IN HeLA CELLS.	247
FIGURE 7.9. LOCALISATION OF GFP AND CT442-GFP IN CT-L2 INFECTED HeLA CELLS.	248
FIGURE 7.10. LOCALISATION OF CALRETICULIN AND CT442-GFP IN HeLA CELLS.	250
FIGURE 7.11. ANTI-CT442 ANTIBODY BINDS CT442 FROM CT-INFECTED HeLA CELLS.	251
FIGURE 7.12. ANTI-CT442 ANTIBODY STAINING OF THE INCLUSION MEMBRANE IN CT-L2 INFECTED HeLA CELLS.	253
FIGURE 7.13. ANTI-CT442 STAINS CT442-GFP IN HeLA CELLS.	254
FIGURE 7.14. ANTI-CT442 ANTIBODY STAINING OF THE INCLUSION MEMBRANE SUSTAINS FROM 24 HPI TO 72 HPI.	255
FIGURE 7.15. CONFIRMATION OF GFP AND CT442-GFP CO-IMMUNOPRECIPITATION.	256
FIGURE 7.16. FILTERING OF PROTEINS IDENTIFIED BY MASS-SPECTROMETRY FROM CO-IMMUNOPRECIPITATION OF GFP OR CT442-GFP.	258
FIGURE 7.17. A VISUALISATION OF PROTEIN-PROTEIN INTERACTIONS IDENTIFIED FROM CT442-GFP CO-IMMUNOPRECIPITATION BY STRING.	272
FIGURE 7.18. A VISUALISATION OF PROTEIN-PROTEIN INTERACTIONS IDENTIFIED FROM CT442-GFP CO-IMMUNOPRECIPITATION BY STRING.	273
FIGURE 7.19. CONFIRMATION OF Rab7 CO-IMMUNOPRECIPITATION WITH CT442-GFP.	275
FIGURE 8.1. EXTRACELLULAR EXPOSURE OF NON-SURFACE CT ANTIGENS.	288
FIGURE 8.2. BLOCKING AND ENHANCING ANTIBODIES.	293
FIGURE 8.3. DECOY ANTIGENS DIVERT ANTIBODY RESPONSES FROM PROTECTIVE ANTIGENS.	296
FIGURE 1. AVERAGE SILHOUETTE WIDTH COMPARISON OF NUMBER OF CLUSTERS FOR CHAPTER 4.	341
FIGURE 2. COMPARISON OF ELISA AND ARRAY RESULTS FOR 90 LONGITUDINAL SERA.	342
FIGURE 3. AVERAGE SILHOUETTE WIDTH COMPARISON OF NUMBER OF CLUSTERS FOR CHAPTER 5.	343
FIGURE 4. INTEGRATED HAPLOTYPE SCORES PRE- AND POST-IMPUTATION.	344
FIGURE 5. COMPARISON OF IHS SCORES BY MINOR ALLELE FREQUENCY (MAF) AND PERCENTAGE MISSED CALLS.	345
FIGURE 6. COMPARISON OF IHS SCORES FROM DIFFERENT SOFTWARE.	346
FIGURE 7. EVIDENCE OF SELECTION IN CT314, CT545, CT695, CT806.	347
FIGURE 8. NUCLEOTIDE ALIGNMENT OF CT442 FROM SEROVAR-REPRESENTATIVE STRAINS.	348
FIGURE 9. ÄKTA SIZE EXCLUSION-BASED CHROMATOGRAMS FOR CT442-HIS.	349
FIGURE 10. LOCALISATION OF IncB-GFP IN HeLA CELLS.	350
FIGURE 11. LOCALISATION OF NUE-GFP IN HeLA CELLS.	351

List of Tables

TABLE 1.1. CT AND HOST FACTORS INVOLVED IN CELL ENTRY AND EARLY INTRACELLULAR EVENTS.	28
TABLE 1.2. SUMMARY OF CT MICRO-ARRAYS SCREENS OF HUMAN SERA.	37
TABLE 1.3. SUMMARY OF FREQUENTLY AND DIFFERENTIALLY RECOGNISED ANTIGENS IN ADULTS WITH TT.	39
TABLE 3.1. PRIMARY ANTIBODIES FOR IMMNOBLOTTING.....	47
TABLE 3.2. MOLECULAR CHAPERONES EXPRESSED BY THE TAKARA CHAPERONE PLASMIDS.	47
TABLE 3.3. TRIALLED CONDITIONS FOR RESTRICTION DIGEST OF THE CT442-HIS INSERT.....	49
TABLE 3.4. TRIALLED CONDITIONS FOR LIGATION OF THE CT442 INSERT INTO THE PET22B(+) PLASMID.	50
TABLE 3.5. TRIALLED DETERGENTS FOR IMPROVED SOLUBILITY OF EXPRESSED PROTEINS.	51
TABLE 3.6. SUMMARY OF HOW SIMILARITIES WITHIN AND BETWEEN CLUSTERS IMPACT THE SILHOUETTE SCORE.	56
TABLE 3.7. DIFFERENT OUTCOME MEASURES CONSIDERED FOR ANALYSIS OF THE LONGITUDINAL STUDY OF Ct INFECTION....	58
TABLE 3.8. SAMPLE CHARACTERISTICS FROM THE HLA POLYMORPHISM AND SCARRING TRACHOMA STUDY.....	62
TABLE 3.9. SAMPLE CHARACTERISTICS FROM THE LONGITUDINAL COHORT STUDY OF Ct INFECTION AND ACTIVE TRACHOMA.	64
TABLE 3.10. SAMPLE CHARACTERISTICS FROM THE 2006-2009 SCARRING CASE-CONTROL STUDY.	65
TABLE 3.11. SAMPLE CHARACTERISTICS FROM THE NK-CELLS AND SCARRING TRACHOMA STUDY.	65
TABLE 3.12. DEMOGRAPHICS OF CHILDREN SAMPLED AT THE END OF A 4-YEAR LONGITUDINAL COHORT OF SCARRING PROGRESSION IN TANZANIA.	66
TABLE 3.13. PRIMARY ANTIBODIES FOR IF MICROSCOPY.	79
TABLE 3.14. GROUPING OF PROTEINS IDENTIFIED BY MASS SPECTROMETRY.	81
TABLE 4.1. PATIENT DEMOGRAPHICS IN PROTECTED AND SUSCEPTIBLE GROUPS.....	88
TABLE 4.2. UNEVEN ANTIBODY RESPONSES IN PROTECTED AND SUSCEPTIBLE INDIVIDUALS, AS DETERMINED BY HILL NUMBERS.	94
TABLE 4.3. FORT-TWO DIFFERENTIALLY RECOGNISED ANTIGENS BETWEEN PROTECTED AND SUSCEPTIBLE INDIVIDUALS.	96
TABLE 4.4. COMBINATORIAL ANTIBODY RESPONSES INCREASED COMPLEXITY WITH MINIMAL BENEFIT TO PREDICTIVE POWER.	101
TABLE 4.5. MOST IMPORTANT ANTIGENS FOR CLASSIFICATION BY RANDOM FORESTS REGRESSION.....	105
TABLE 4.6. FIVE ANTIGENS ASSOCIATED WITH PROTECTION FROM INFECTION.	108
TABLE 4.7. PATIENT DEMOGRAPHICS IN PROTECTED AND SUSCEPTIBLE GROUPS, COMPLETE SERA SET.	121
TABLE 4.8. AGREEMENT OF CONCORDANT AND DISCONCORDANT SERA FROM CT089 AND CT875.....	127
TABLE 4.9. ELISA AND ARRAY RESULTS AFTER EXCLUDING "BOTH DISCORDANT" SERA.	127
TABLE 4.10. ELISA RESULTS FROM THE COMPLETE 130 SERA.....	130
TABLE 5.1. PATIENT DEMOGRAPHICS IN PROTECTED AND SUSCEPTIBLE GROUPS.....	145
TABLE 5.2. UNEVEN ANTIBODY RESPONSES IN ADULTS WITH AND WITHOUT SCARRING, AS DETERMINED BY HILL NUMBERS.	152
TABLE 5.3. DIFFERENTIALLY RECOGNISED ANTIGENS BETWEEN ADULTS WITH AND WITHOUT SCARRING.	154
TABLE 5.4. COMBINATORIAL ANTIBODY RESPONSES INCREASED PREDICTIVE POWER.	156
TABLE 5.5. MOST IMPORTANT ANTIGENS FOR CLASSIFICATION BY RANDOM FORESTS REGRESSION.....	158
TABLE 5.6. PATIENT DEMOGRAPHICS IN ADULTS WITH AND WITHOUT SCARRING, COMPLETE SET OF 116 SERA.	168
TABLE 5.7. ELISA RESULTS FROM THE COMPLETE 116 SERA AND TWO FURTHER SCARRING CASE-CONTROL STUDIES.....	174

TABLE 5.8. DEMOGRAPHICS OF CHILDREN SAMPLED AT THE END OF A 4-YEAR LONGITUDINAL COHORT OF SCARRING	
PROGRESSION IN TANZANIA.	175
TABLE 5.9. ELISA RESULTS FROM THE 311 CHILDREN FOLLOWED AS PART OF A 4-YEAR LONGITUDINAL STUDY OF SCARRING	
PROGRESSION IN TANZANIA.	177
TABLE 6.1: EVIDENCE OF NATURAL SELECTION BY COMBINING TAJIMA'S D, FU AND LI'S D*/F* AND FAY AND WU'S H....	188
TABLE 6.2. PREDICTIONS OF IMMUNOGENIC EPITOPES, TRANSMEMBRANE DOMAINS AND CELLULAR LOCALISATION OF IMMUNE	
TARGETS FROM CHAPTERS 4 AND 5	192
TABLE 6.3. GENES UNDER SELECTION IDENTIFIED BY TAJIMA'S D.	201
TABLE 6.4. GENES UNDER BALANCING SELECTION BY TAJIMA'S D.	202
TABLE 6.5. GENES UNDER SELECTION IDENTIFIED BY TAJIMA'S D AND FAY AND WU'S H.	205
TABLE 6.6. GENES WITH SLIDING WINDOW-LEVEL EVIDENCE OF SELECTION BY TAJIMA'S D AND FAY AND WU'S H.	206
TABLE 6.7. REGIONS UNDER POSITIVE SELECTION IDENTIFIED BY IHS.	209
TABLE 6.8. EVIDENCE OF SELECTION IN IMMUNE TARGETS IDENTIFIED IN CHAPTERS 4 AND 5.	217
TABLE 7.1. GROUPING OF PROTEINS IDENTIFIED BY MASS SPECTROMETRY.	257
TABLE 7.2. SUMMARY OF NUMBER OF PROTEINS IDENTIFIED AFTER FILTERING.	258
TABLE 7.3. DETAILS OF FILTERED PROTEINS WITH A MASS SPECTROMETRY SCORE GREATER THAN 50.	260
TABLE 7.4. GO-TERM PROCESS AND FUNCTION ENRICHMENT OF RAB PROTEINS AND PROTEIN TRANSPORT.	266
TABLE 7.5. PROTEIN DOMAIN ENRICHMENT OF GTPASES.	267
TABLE 7.6. PATHWAY ANALYSIS ENRICHMENT OF METABOLIC PATHWAYS AND PROCESSES RELATED TO INTRACELLULAR	
TRAFFICKING.	268
TABLE 7.7. RANKED PROTEIN-PROTEIN INTERACTIONS IDENTIFIED BY STRING.	269
TABLE 7.8. CT PROTEINS IDENTIFIED THROUGH CT442-GFP CO-IMMUNOPRECIPITATION.	274
TABLE 0.1. SUMMARY OF PREVIOUS CT MICRO-ARRAY ANTIGEN IDENTIFICATION.	336
TABLE 0.2. CT442-HIS PRIMERS FOR CLONING INTO PET22(B)+.	340

Abbreviations

AIC: Akaike information criterion

AUC: Area under the curve

CD: Cluster of differentiation

CI: Confidence intervals

CO: Corneal opacity

Ct: *Chlamydia trachomatis*

EB: Elementary body

EHH: Extended haplotype homozygosity

ELISA: Enzyme-linked immunosorbent assay

GFP: Green fluorescence protein

Glm: Generalised linear model

GST: Glutathione S-transferase

His: Histidine

HPI: Hours post infection

HSP: Heat shock protein

IDO: Indoleamine 2,3-dioxygenase

IFN: Interferon

Ig: Immunoglobulin

iHS: Integrated haplotype score

IL: Interleukin

Inc: Inclusion membrane protein

IPTG: Isopropyl β -D-1-thiogalactopyranoside

IWR: Interquartile range

LGV: Lymphogranuloma venereum

MAF: Minor allele frequency

MDA: Mass drug administration

MIP: Macrophage infectivity potentiator

MOMP: Major outer membrane protein

NK: Natural killer

NUE: Nuclear effector

OD: Optical density

OR: Odds ratio

PFA: Paraformaldehyde

Pmp: Polymorphic membrane protein

RB: Reticulate body

ROC: Receiver operating characteristic

RPM: Revolutions per minute

RRD: Relative rank deviation

SNP: Single nucleotide polymorphism

TARP: Translocated actin-recruiting phosphoprotein

TF: Trachomatous inflammation-follicular

TFI: Tubal factor infertility

TI: Trachomatous inflammation-intense

TS: Trachomatous scarring

TT: Trachomatous trichiasis

1. Introduction

1.1. Trachoma

1.1.1. Epidemiology

Trachoma, caused by ocular infection with *Chlamydia trachomatis* (Ct), is the leading infectious cause of blindness worldwide. Ocular infections with Ct affect the epithelial cells of the conjunctiva¹, with repeated infection in endemic areas causing a chronic keratoconjunctivitis^{2, 3}. While numbers of people blinded by trachoma have declined in the past two decades, from over 6 million in the 1990's⁴ to a present day estimate of around 1.1 million⁵, there is still a significant disease burden. At least 2.8 million people suffer some form of visual impairment⁶, with an estimated 40 million people having active disease, together accounting for a loss of 1 million disability-adjusted life years and substantially reduced productivity⁷.

Ct infection causes an influx of innate and adaptive immune cells into the conjunctiva, including neutrophils, NK cells, macrophages, T-cells and B-cells. B-cells are the predominant cell type which form observed conjunctival follicles, known as trachoma inflammation, follicular (TF). Higher load or persistent infections cause a more severe inflammatory infiltration, indicated by vascularisation/reddening of the conjunctival surface, known as trachoma inflammation, intense (TI). Repeated rounds of infection and subsequent bouts of inflammation cause damage to the conjunctival tissue which results in deposition of scar tissue as part of the healing process, known as trachomatous scarring (TS), however as this scarring progresses over time, the eyelids become distorted leading to in-turned eyelids and eyelashes, known as entropion. In-turned eyelashes in contact with the surface of the eye, known as trachomatous trichiasis (TT), cause mechanical abrasions which can lead to corneal opacities (CO)^{1, 8} (Figure 1.1).

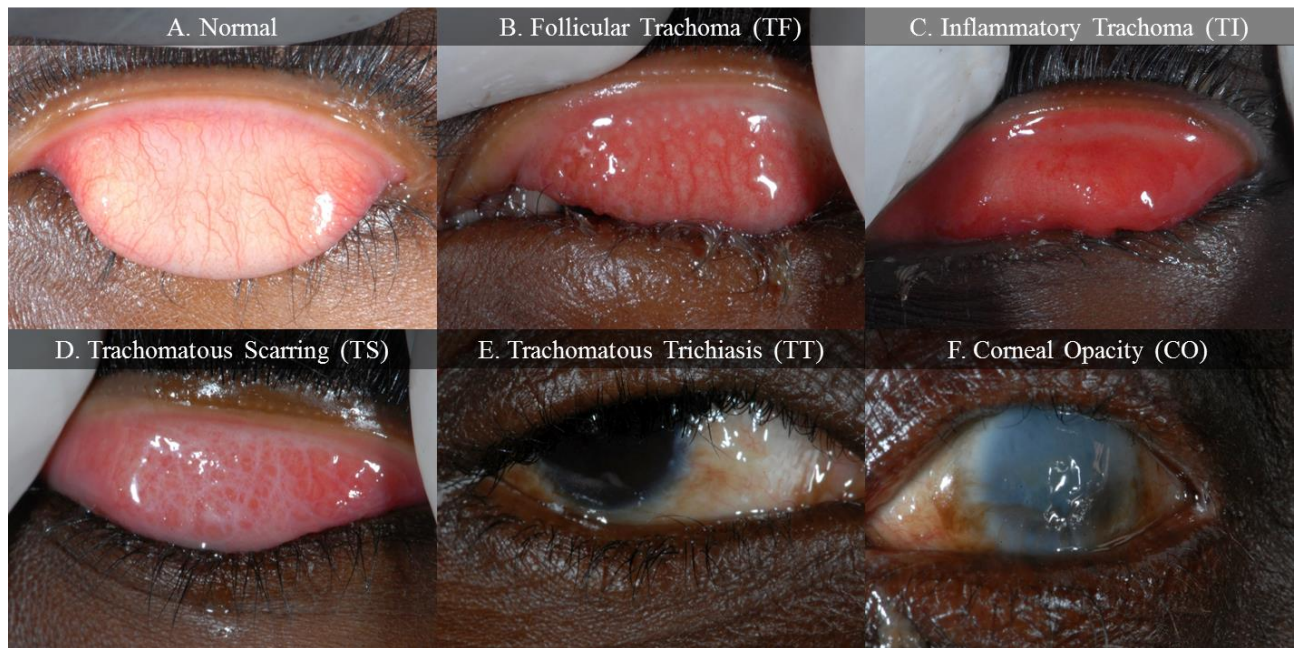


Figure 1.1. WHO simplified trachoma grading system.

Infections causing the clinical signs of active trachoma are almost exclusive to children, who harbour both the infection and disease for longer but suffer minor scarring pathology^{8, 9}. In contrast infection and active disease are rarely observed in adults, but scarring pathology is more frequently observed with the highest prevalence of scarring disease observed in people over 40 years of age. The prevalence of conjunctival scarring and visual impairment varies between trachoma endemic populations, areas with the highest prevalence of active disease also have the highest prevalence of scarring and blinding disease.

Currently trachoma is considered to be endemic in 55 countries, with recent trachoma assessment data available from 31 of these. Trachoma is widespread, being found in Central and South America, South and Southeast Asia, Australia and the Pacific Islands¹⁰. However it is most common in dry regions of Africa, notably in East and Central Africa and the Sahel region of West Africa (Figure 1.2)^{10, 11}. Trachoma control efforts are underway in many of these areas, and several countries have eliminated trachoma as a public health problem¹². Despite this many countries still lack reliable data and are yet to implement control measures^{11, 13}.

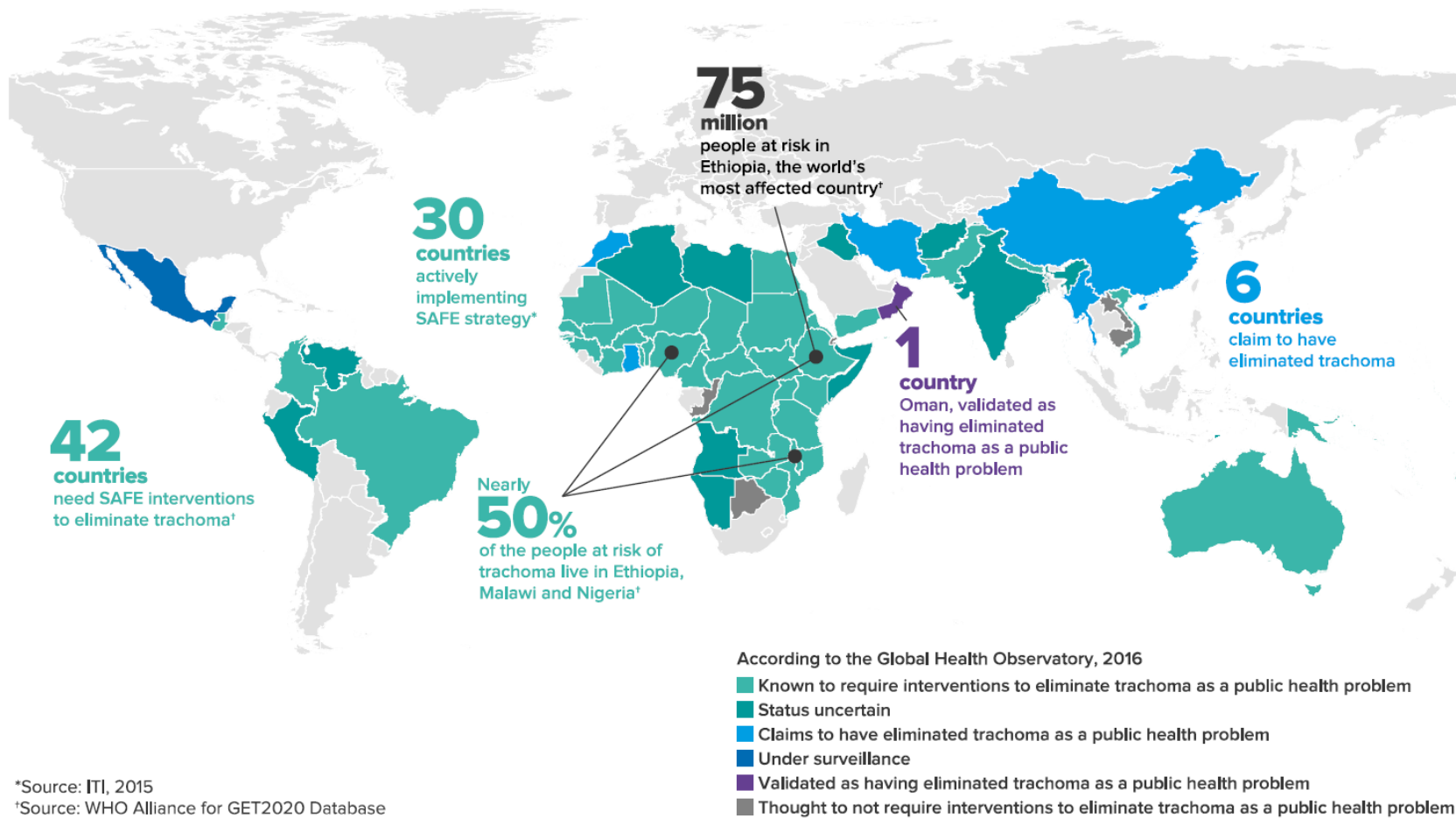


Figure 1.2. Global map of active trachoma (2015).

Image courtesy of the International Coalition for Trachoma control (<http://www.trachomacoalition.org/>).

1.1.2. Control and elimination

Global Elimination of Trachoma by 2020 (GET2020) is a WHO-endorsed alliance focussed on increased and improved mapping of endemic areas and control through the SAFE strategy^{14, 15}. The strategy consists of Surgery for in-turned eyelids, Antibiotics for infection, Facial cleanliness to reduce transmission and Environmental change for improved sanitation to reduce transmission (Figure 1.3). The antibiotic of choice for trachoma is the Pfizer-donated Zithromax, administered annually in endemic communities depending on the prevalence of clinical signs of active disease in the 1-9 year olds, with surveys 3-5 years after implementation to decide on continued annual treatment and/or reinforcement of FE components¹⁶⁻¹⁸.



Figure 1.3. SAFE strategy for trachoma control.

1.1.3. Risk factors for trachoma

Risk factors for trachoma have been extensively reviewed¹⁹⁻²¹. Establishing causal relationships between individual factors and trachoma has been challenging due to their complex and interconnected nature. Risks include environmental factors, socio-economic status, behavioural factors, host genetics and immune response.

Trachoma is a seasonal disease²² and is predominantly associated with dry areas, most commonly in sub-Saharan Africa. Availability of water and type of water source

are known risk factors^{23, 24}. These are linked with the universal finding that facial cleanliness^{19, 20, 23}, individually and in combination with ocular/nasal secretions, is associated with trachoma. The presence of flies (*Musca sorbens*), either on the faces of children or around latrines, have an inconsistent association^{25, 26}. It is unclear whether they are able to transmit Ct or are just a consequence of poor facial hygiene. Similarly, the presence of other pathogens and commensal bacteria on the conjunctiva is associated with trachoma²⁷, this may be both a result of trachomatous disease and also a driving factor.

Socio-economic status is an important risk factor for trachoma. Crowded households are associated with increased risk of trachoma²⁰, particularly in terms of numbers of children as they are at greater risk.^{24, 28} Lower levels of household education are also a risk factor²⁹. The presence of household cattle increases the risk of trachoma^{21, 23}, this may be linked to the presence of flies.

1.2. *Chlamydia trachomatis*

1.2.1. History and evolution within and between species

The causative organism Ct belongs to the *Chlamydiaceae* family of Gram-negative-like, obligate, intracellular bacteria, which contains two separate genera *Chlamydia* and *Chlamydophila*³⁰. However the evolution of the species within these genera, and their closely related biology and ecology, suggest they should more appropriately be considered as a single genus³¹. The family is believed to have arisen around 200 million years before present^{32, 33}, and during this process they lost large parts of their genome in adaptation to intracellular life retaining only necessary genes³⁴. This evolutionary optimisation of the genomes is supported by expression of over 90% of their genes at the mid-point of their life cycle³⁵. *Chlamydia trachomatis* as a separate species evolved around 100 million years ago with the early mammals^{31, 33, 36}.

Ct can be split into two biovars based on tissue tropism and other biological properties, lymphogranuloma venereum (LGV) which may disseminate and spread to the lymphatics and trachoma biovars which are more restricted to the urogenital tract or conjunctiva. The trachoma biovar can be further subdivided based on serological typing

of the major outer membrane protein (MOMP, *ompA* gene). Ocular and urogenital strains are subdivided into 16-19 serovars designated A to K. A, B and C serovars are predominately associated with ocular infection and D-K serotypes with urogenital infection. Phylogeny based on the results of whole genome sequencing, rather than *ompA* genotyping, suggests that in evolutionary terms the A-C genotypes diverged from the urogenital strains around 2.5 million years before present³² after the trachoma biovar had previously split into the groups T1 and T2³⁷ (Figure 1.4, adapted from Harris *et al*³⁷). A recent study of conjunctival Ct isolates from an Australian Aboriginal community have questioned the monophyletic lineage of ocular isolates, they found ocular isolates more closely related to both the T1 and T2 urogenital lineages except for apparently ocular *ompA* and certain polymorphic membrane protein (*pmp*) genes³⁸. The evolutionary origin of these unusual isolates and the evidence for trachoma in the sampled children require further study.

Ct serovars A-C have been consistently isolated from the ocular conjunctiva of individuals with trachoma³⁹⁻⁴¹. Studies in the 1950s and 1960s demonstrated Ct was able to be cultured *in vitro*⁴². Subsequent inoculation of blind volunteers confirmed Ct as the causative agent of trachomatous disease⁴³⁻⁴⁵.

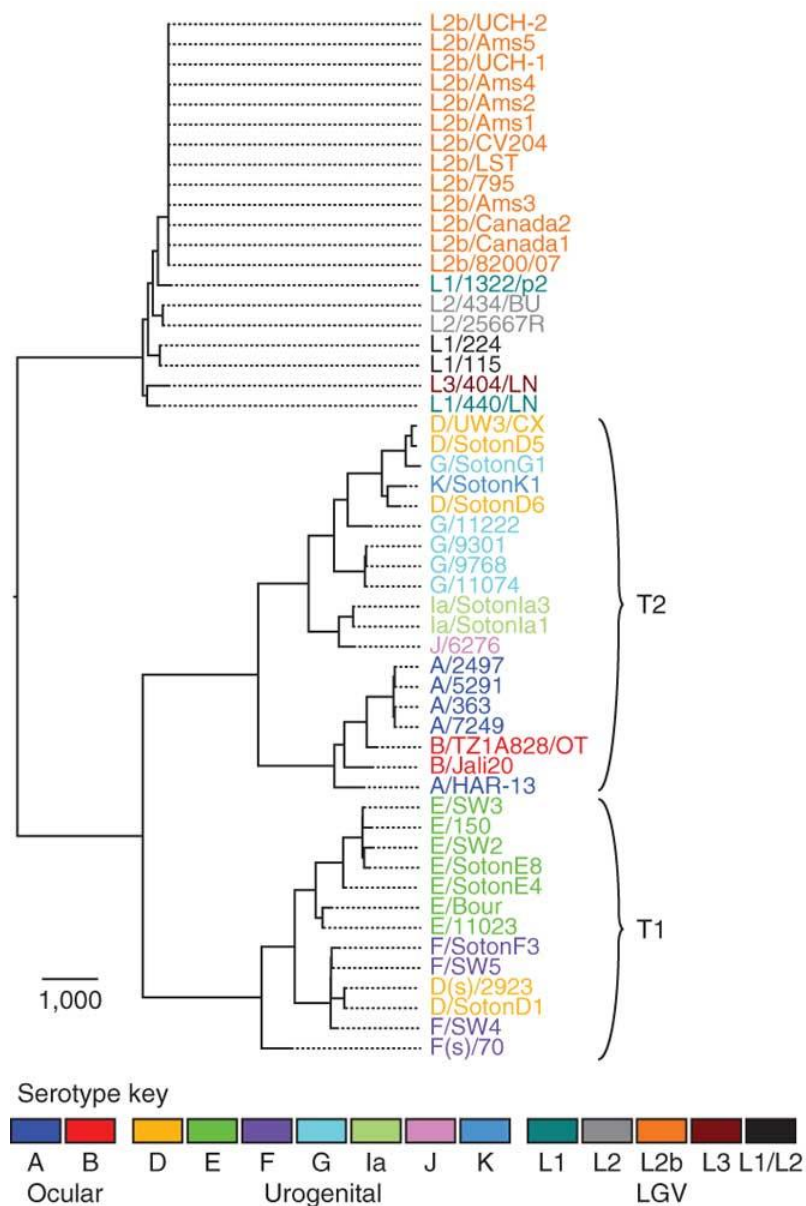


Figure 1.4. Maximum likelihood reconstruction of the Ct whole-genome phylogeny.

Ct species phylogeny using the chromosomal sequences of 52 genomes. Strain names are coloured by serotype, see key. Scale bar represents number of SNPs.

1.2.2. Host cell entry and early host interactions

The developmental cycle of Ct has been defined extensively *in vitro*. During the cycle Ct alternates between an extracellular, stress-resistant infectious form (elementary body) and an intracellular, non-infectious replicative form (reticulate body). Differentiation from elementary bodies (EBs) to reticulate bodies (RBs) occurs intracellularly after invasion of host cells, following several rounds of replication RBs differentiate back to EBs for reinfection of new host cells (Figure 1.5, adapted from Bastidas *et al*⁴⁶).

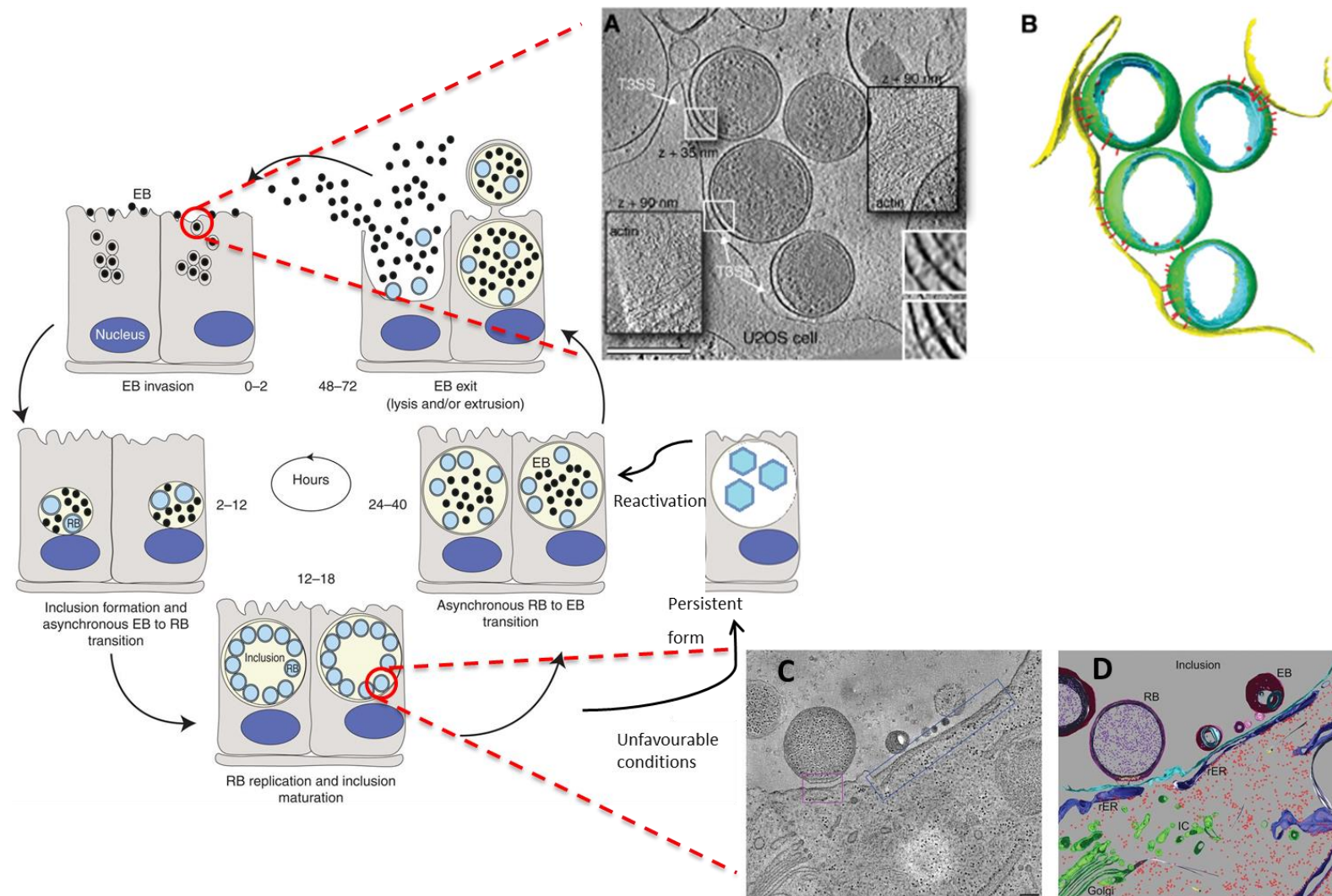


Figure 1.5. Ct developmental cycle.

A) Cryo-electron microscopy tomogram of EBs contacting the host plasma membrane through the type-three secretion (T3S) apparatus. B) 3D representation. C) Electron microscopy tomogram of the 'synapse' across the inclusion membrane involving the RER, T3S-apparatus and RBs. D) Labelled-electron tomography data.

Ct enters cells as metabolically-inactive, infectious EBs, these are resistant to stress owing to a rigid outer membrane supported by disulphide cross-linked cysteine-rich proteins^{47, 48}. Prior to contacting host cells, EBs polarise with the bacterially-conserved type-three secretion (T3S) machinery⁴⁹ aggregating on the side of host-cell contact⁵⁰ (Figure 1.5 parts A and B, adapted from Nans *et al*⁵⁰). Preliminary low affinity interactions with host heparan sulphate proteoglycans⁴⁶ are quickly strengthened by higher affinity binding involving various chlamydial and host proteins (Table 1.1). The best characterised of the chlamydial-entry proteins is TARP (translocated actin-recruiting phosphoprotein), a T3S-effector which enters the host cytosol upon contact with the plasma membrane^{47, 51}. Through interactions with SH2-domain containing proteins Tarp can induce cellular signalling and actin polymerisation and reorganisation^{48, 52}, directly and through Arp2/3. Through SHC1 TARP can also inhibit apoptosis⁵³, the importance of this is supported by infection-related upregulation of anti-apoptotic Bag1 through MAPK/ERK signalling⁵⁴ and transcriptional profiling of mRNA³⁵ and microRNA⁵⁵ data showing modulation of apoptosis-related transcripts. Another effector TepP has been shown to act immediately downstream of Tarp, where it recruits scaffold proteins to impact host signalling and innate immune responses⁵⁶. These early events are critical in entry and survival⁵³. Another T3S-effector, CT694, is believed to be important in reversing these cytoskeletal changes once the bacteria have entered⁴⁶.

The next step in intracellular survival is avoidance of lysosomal degradation, for which Ct employs two strategies. Initially EBs delay the maturation of their encapsulating vacuoles to avoid fusion with lysosomes, following this they form an intracellular compartment, the inclusion, which actively modulates host interactions and responses⁵⁷.

Table 1.1. Ct and host factors involved in cell entry and early intracellular events.

CHLAMYDIAL PROTEIN	HOST PROTEIN(S)	FUNCTION(S)	REFERENCE(S)
UNKNOWN	FGF2 (Fibroblast growth factor 2)	Enhance EB binding and uptake, heparan sulphate proteoglycan-	58, 59

		dependent, via FGF-receptor	
UNKNOWN	FGFR (Fibroblast growth factor receptor)/PDGFRb (Platelet-derived growth factor receptor b)	Stimulate mitogenic signals through Erk1/2, latter involved in entry	52, 60
UNKNOWN	Dynamin2, Hsc70 and Arp2	Clathrin-mediated endocytosis and interaction with Tarp-mediated entry	61
LPS	CFTR (Cystic fibrosis transmembrane conductance regulator)	Required for entry and uptake of EBs, impairs the normal ion-channel function of CFTR	62
TARP	Various	Regulates apoptosis, cell growth and intracellular signalling	53
TEPP	Crkl-2 (Scaffold protein)	Downstream of Tarp, recruits scaffold proteins upon entry, involved in signalling an innate immune response	56
CT694	AHNAK and others	Localises to the plasma membrane, possibly reverses Tarp-actin reorganisation	46, 63, 64
PMPD	Unknown	Important in attachment and	65, 66

		stimulation
MOMP	Possibly mannose-6 phosphate	Role in attachment, via heparan-sulphates, but not necessary ⁶⁷
OMCB	Unknown	Role in attachment, via glycosaminoglycan-like receptors ⁶⁸

1.2.3. Chlamydial inclusion and its host interactions

The early inclusion consists of host proteins derived from its vesicular carriers, as it develops and chlamydial protein synthesis starts these replace those of the host⁴⁸. Inclusions are found at the microtubule-organising centre, owing to their early vesicle trafficking, in close association with the nucleus⁶⁹. They are stabilised by interactions with the host cytoskeleton⁴⁸ and this localisation allows association with the Golgi apparatus, which in acute but not persistent infection⁷⁰ is actively fragmented to form Golgi stacks for host interactions⁷¹. Within this compartment EBs transform into metabolically-active RBs which replicate through binary fission⁴⁸. The inclusion protects the bacteria from host attack and facilitates nutrient acquisition and host modulation. Nutrients are acquired through selective sequestering of lipid-containing vesicles in the endosomal pathway⁷², additionally lipid droplets can enter the inclusion^{48, 73} and membrane-associated peroxisomes can be taken up⁷⁴.

Host modulation is achieved through secretion of chlamydial effectors⁴⁶, which involves direct interaction with the rough endoplasmic reticulum (RER) (Figure 1.5 parts C and D, adapted from Dumoux *et al* ⁷⁵)⁷⁶. Many of these effectors are secreted into the host cytosol by the type-2 secretion (T2S) or T3S-apparatus^{77, 78}, including Tarp, CT694 and CPAF (chlamydial proteasome like activity factor)⁷⁹⁻⁸¹. Additionally, a number of effectors that can enter the host nucleus such as NUE (nuclear effector) have been identified⁸². Between 50 and 65 of the T3S-proteins after entering the cytosol are inserted into the inclusion membrane, by a currently unclear mechanism. These inclusion membrane proteins (Incs), while dissimilar in primary sequence, share a bi-lobed membrane-spanning structure and interact extensively with the host cell⁸³.

Host organelles including Golgi-derived vesicles⁸⁴, multi-vesicular bodies⁸⁵, lipid droplets⁸⁶, rough endoplasmic reticulum⁷⁶ and centrosomes⁸⁷ are recruited to the inclusion and are important in its development, acquisition of nutrients and Ct survival and transmission. Until recently only a few Incs had been functionally investigated. CT223, CT229 and IncA are important in inclusion trafficking and development. CT223 recruits the host factor CEP170 through which it modulates microtubule organisation to traffic towards centrosomes⁸⁷. IncA is essential for homotypic fusion of inclusions, the process by which inclusions fuse in multiply infected cells⁸⁸. CT229 recruits Rab4 as early as 2 hours post-infection, likely functioning in trafficking of the inclusion and recruited host organelles⁸⁹. A number of other Rab proteins are recruited to the inclusion⁹⁰. As well as intracellular trafficking these Rab proteins are critical for lipid acquisition in Ct, which requires Golgi disruption as described previously⁹¹. IncD through the lipid transfer protein CERT⁹² and IncG through recruitment of lipid droplets⁹³, are also important in lipid acquisition. IncG also recruits the host adaptor protein 14-3-3 β through which it regulates Ct-resistance to apoptosis^{94, 95}.

More recently a large-scale affinity-purification mass spectrometry analysis was performed on Incs, identifying host interactions for 38 Incs⁹⁶. Several common functional pathways were identified including; endocytosis, ubiquitination, apoptosis, cell division and DNA damage/repair. This study identified an interaction between IncE and two components of the retromer, a complex of host proteins involved in retrograde trafficking from endosomes to the trans-Golgi network. Recruitment of sorting nexins 5 and 6 to the inclusion prevented association of the retromer complex, inhibition of this function significantly reduced Ct progeny formation. This approach did not identify all previously identified Inc interactors, but it did highlight organelles and pathways either known to or assumed to be important during Ct intracellular development.

Incs also regulate Ct exit from cells. Many Incs are localised uniformly throughout the inclusion membrane, a subset localises to discrete microdomains⁹⁷. These microdomains are enriched for Src family kinases, suggesting they may control multiple functions through downstream signalling. CT850 interacts with dynein motor proteins to regulate inclusion intracellular positioning⁹⁸, another microdomain Inc CT228 appears to control cell exit⁹⁹. Through differential recruitment and activation of components of the myosin phosphatase pathway, CT228 can regulate the switch between two methods of exit described below.

1.2.4. Exit and the potential for persistence

In vitro between 24 and 48 hours after infection, once the RBs have replicated extensively, the reversion back to EBs begins. This brings about changes in gene expression, with a focus on gene involved in modifying the bacterial surface, amino acid synthesis and energy metabolism¹⁰⁰. Around 60 hours post-infection release of the infectious EBs can be seen¹⁰⁰. Previously it was thought to occur only through cell lysis, where the inclusion and subsequently the host plasma membrane rupture to release infectious EBs^{46, 72}. It is now known that packaged release involving extrusion of the inclusion towards the plasma membrane and cytoskeletal reorganisation is also possible^{46, 72}. Extrusion is an active process involving actin cytoskeletal reorganisation through which EBs pinch off from the inclusion, surrounded by inclusion membrane, followed by protrusion and detachment from the host cell membrane⁷². This leaves the host cell intact and often residual inclusion remains. Recent studies have suggested extrusions can enhance extracellular survival and viability of EBs, while also facilitating uptake and survival within macrophages¹⁰¹.

An interesting phenomenon of *Chlamydia* species is their *in vitro* ability to ‘persist’ in a dormant, morphologically abnormal¹⁰²⁻¹⁰⁴, yet viable state¹⁰⁵. It can be induced by antibiotics, iron depletion¹⁰³, amino acid starvation, oxidative stress and IFN γ ¹⁰⁶. IFN γ -induction depletes intracellular tryptophan causing upregulation of tryptophan synthesis and DNA repair pathways, while reducing cell division and RB-EB transition, supporting a persistent phenotype¹⁰⁵. The *in vivo* existence of persistence has not been proven.

1.2.5. *Chlamydia trachomatis* population genetics

Prior to the availability of Ct whole-genome sequences, Ct genetics was focussed on a few key genes involved in pathogenicity and tropism. *OmpA* as described above was used to define the serovars, it was also commonly used to investigate transmission in trachoma-endemic communities. A study in The Gambia found minimal change in the prevalence of individual *ompA* variants or the sequence of these variants during a 2-year study¹⁰⁷. A later study in The Gambia had similar results examining *ompA* alleles collected before and 2 months after mass drug administration (MDA) of Zithromax in a

community. There was evidence of directional selection, with most new mutations being rare either because an allele had risen to fixation (positive selection) or variation was being driven out of the population (purifying selection)¹⁰⁸. A study of *ompA* in Tanzania found evidence of positive selection in serovar B isolates and purifying selection in serovar A isolates¹⁰⁹, suggesting natural selection was acting differently within these two Ct populations.

OmpA does not reflect the whole-genome phylogeny described previously, however the plasticity zone, *tarP* and the polymorphic membrane proteins (*pmp*) do associate with tropism. The plasticity zone is a highly variable region of approximately 25 genes in Ct, several of which are pseudogenes within certain serovars. Polymorphisms within this region, notably in a putative cytotoxin and the tryptophan synthase operon, distinguish between the tropisms¹¹⁰. In ocular isolates truncation of *trpA* or mutations in *trpB* mean they cannot synthesise tryptophan¹¹¹. During Ct infection high levels of IFN γ induce IDO which promotes degradation of tryptophan. Indole from the urogenital microbial flora is hypothesised to allow urogenital and LGV isolates to synthesise tryptophan, this inability to synthesise tryptophan is thought to be part of the reason serovar A-C isolates are primarily restricted to the ocular conjunctiva. The domain structure of the Ct-entry effector TARP varies between the ocular, urogenital and LGV biovars¹¹². Numbers of tyrosine-rich repeats and actin-binding domains differentiate between the three groups, possibly affecting invasive ability. *PmpF*, *pmpH* and to a lesser extent *pmpB*, *C*, *G* and *I* associate with tropism¹¹³. *PmpB* had evidence of positive selection across 18 Ct isolates, suggesting it is important in Ct survival. All remaining *pmp*'s were under purifying selection or evolving neutrally.

Genome-wide scans for genes under selection have consistently supported these genes as being important in Ct pathogenicity and tropism. Using populations of Ct isolates three studies have been published using 4¹¹⁴, 12¹¹⁵ and 59¹¹³ genomes respectively. The first two studies identified *tarP*, *ompA* and *pmp*'s under positive selection across all studied serovars, they also identified a number of Inc's under positive selection. The study using 59 genomes tested for evidence of positive selection associated with different aspects of Ct evolution. Adaptation to the conjunctiva and variations in ocular pathogenicity highlighted the same genes again, further demonstrating their importance in Ct virulence. A study of 4 ocular isolates with differential virulence in non-human primates and *in vitro* found remarkably similar polymorphic genes, including *tarP*, 3 genes in the plasticity zone and CT147, the Inc

putatively involved in Ct avoidance of lysosome fusion¹¹⁶. All of these studies also found genes classed as hypothetical under positive selection within Ct, highlighting the considerable portion of the genome currently not characterised.

1.3. Immunology of trachoma

1.3.1. Initial responses to Ct infection

Ct infection of conjunctival epithelium induces a strong pro-inflammatory response, characterised by production of cytokines and chemokines such as interleukin-1 alpha (IL-1 α), IL-6, IL-8 and granulocyte macrophage colony-stimulating factor (GMCSF)^{117, 118}. This response stimulates infiltration of neutrophils, macrophages and NK cells¹¹⁹⁻¹²¹.

This innate immune response stimulates an influx of adaptive immune cells. Conjunctival follicles initially primarily consist of B cells, with evidence of macrophages and T-cells¹¹⁹. This is supported by an excess of plasma cells directly below the conjunctiva^{3, 122}. Epithelial cells upregulate MHC-Class 1 and 2 on their surface^{119, 123}. Lymphoproliferative and cytolytic responses to Ct antigens including whole EBs and MOMP are important in the resolution of infection¹²⁴⁻¹²⁶.

Production of IFN γ is an important part of this response^{125, 127}, levels of the cytokine are boosted during Ct infection and active disease. IFN γ is believed to help control ocular infection through induction of IDO and subsequent degradation of tryptophan, as described previously¹¹¹. However, studies of longitudinal infection in children from trachoma-endemic communities found higher levels of IFN γ , and also IDO and IL-10, was associated with longer duration of infection and quicker acquisition of infection¹²⁸.

Local and serum antibody responses are induced by Ct^{129, 130}, their role in immunity is unclear. Anti-Ct IgG, both in tears and serum, associates with active disease with titre increasing with severity of inflammation¹³⁰. However, Ct-infectivity neutralising antibodies have been demonstrated *in vitro* and in animal models¹³¹⁻¹³⁵. Additionally, a strong antibody recall response was associated with partial immunity to reinfection in non-human primates after immunisation with an attenuated plasmid-free Ct¹³⁶. The lack of a consistent immune response associating with protection from Ct

infection and active disease, and the persistence of conjunctival inflammation in the absence of Ct, have led to the suggestion that a chronic, dysregulated immune response is the driving factor in trachomatous sequelae.

1.3.2. Immunopathology of recurrent Ct infection and persistent inflammation

Historically, explanations for trachomatous pathology have been split into two broad groups, the ‘cellular paradigm’ and the ‘immunological paradigm’¹³⁷. The cellular paradigm states that the innate response from epithelial cells to Ct infection drives pathology. The immunological paradigm states that cell-mediated immunity is the driving factor.

The immunological paradigm was originally focussed around a proposed delayed-type hypersensitivity (DTH) reaction against the Ct antigen HSP60, the hypothesis being that re-exposure to HSP60 can stimulate memory T-cells and drive inflammation and tissue damage through secretion of IFN γ and activation of macrophages. Subsequent studies of HSP60 have been inconsistent. Anti-HSP60 antibody responses in some populations have been associated with scarring and urogenital pathology¹³⁸⁻¹⁴⁰, however lymphoproliferative responses appear to associate with resolution of Ct infection and are depressed in adults with scarring. In the latter study HSP60 stimulated IL-4 production in scarred adults¹⁴¹, suggesting weakened Th1-type responses could be driving pathology. Evidence from conjunctival transcriptomics have not supported this, instead they have shown the involvement of regulatory T-cells and Th17 cells in active trachoma^{142, 143}.

Human studies of the conjunctival transcriptome have supported the cellular paradigm. Pro-inflammatory factors such as psoriasin, CXCL5, TNF α and IL-1 β are upregulated in active disease and scarring^{143, 144}. Factors involved in tissue remodelling and fibrosis such as CTGF and several matrix metalloproteinases (MMPs) are also upregulated in scarring^{144, 145}. The exacerbation of conjunctival inflammation by non-chlamydial bacteria, both pathogenic and commensal, further supports the role of innate responses in driving scarring pathology^{27, 146}. In natural infections there is likely a careful balance between cellular and immunological responses, the dysregulation of either may support the chronic inflammatory environment that drives conjunctival scarring pathology.

1.4. Immune targets in Ct infection

1.4.1. Early studies of immunodominant antigens

Initial identification of immunodominant antigens in ocular Ct infection utilised immunoblotting. EB proteins from ocular serovars were separated by gel electrophoresis and blotted with tears and/or serum from individuals with trachoma or previously challenged non-human primates. Antibodies against MOMP, chlamydial LPS and HSP60 were frequently identified in these studies¹⁴⁷⁻¹⁴⁹. Immunoblotting highlighted the presence of multiple unidentified protein bands with weaker recognition, these were stimulating antibody responses but not as frequently or as strongly as MOMP. A recent study in non-human primates immunoblotted with serum after challenge with an attenuated plasmid-free Ct¹³⁶. They identified the immunodominant antigens described above and several other antigens. Based on previous studies they predicted and proved that these unidentified antigens included CPAF, Pgp3 and PmpD. A number of antigens still were unidentified. These studies were limited by the need for antibodies specific to each antigen to confirm their identity, this inherently required previous knowledge of them being immunogenic.

1.4.2. The history and importance of protein-array screens for serological profiling

The last decade has seen the advent of protein-based screens of human serum to discover the complete profile of antibody responses stimulated by an infection¹⁵⁰. This has streamlined the identification of diagnostic and vaccine candidates, leading to faster progression and evaluation of individual targets. For *Plasmodium falciparum* and malaria the proteome has been screened to find immunity-associated antigens¹⁵¹ and antigens associated with particular stages of infection¹⁵², subsequent work used candidates from these studies to test for changes in antibody dynamics with age¹⁵³. Similarly in studies of schistosomiasis serological-screens have been developed¹⁵⁴ and utilised for testing association of antigens, particular antibody isotypes and disease intensity¹⁵⁵. This approach has been applied to a number of bacterial species including *Mycobacterium tuberculosis*. An array was developed using 4000 proteins (99% of the genome) and used to determine antibody profiles of people with and without disease¹⁵⁶.

Candidates from this screen have been used in studies of *in vivo* protection¹⁵⁷, for diagnostics^{158, 159} and to test antibody dynamics in different disease severities and outcomes¹⁶⁰.

1.4.3. Protein-array screens of serological and cellular responses from persons with Ct infection

There have been 6 studies which have screened human serum against micro-arrays of Ct antigens to identify serological responses, 2 of these simultaneously investigated T-cell responses (Table 1.2). Four of these 6 studies utilised the same GST-fusion protein array. One study used a similar array, but utilising both GST and His-tagged proteins. The remaining study produced a genomic expression library of Ct proteins, they identified individual antigens after screening for antibody responses and sequencing the reactive clones. When comparing within these urogenital arrays (Appendix Table 1), there are a number of commonly recognised antigens but there are still many that are identified in only one or two studies. This likely represents differences in production and testing of the arrays and heterogeneity in immune responses, specifically targeted antigens, in Ct infection.

Table 1.2. Summary of Ct micro-arrays screens of human sera.

Number of proteins tested (No. proteins), the characteristics of screened sera (Sera phenotype) and testing for T-cell responses are indicated. UGT CT⁺ means current urogenital Ct infection. TFI means tubal-factor infertility. CT-AB⁺ means tested positive for Ct serum antibodies.

AUTHOR (YEAR)	ARRAY TYPE	NO. PROTEINS	SERA PHENOTYPE	T-CELL RESPONSES
SHARMA (2006)	FUSION PROTEINS	156	FEMALE, UGT CT ⁺	NO
COLER (2009)	EXPRESSION LIBRARY	NA	UGT CT ⁺	YES
WANG (2010)	FUSION PROTEINS	908	FEMALE, UGT CT ⁺	NO

FINCO (2011)	FUSION PROTEINS	120	CT-AB ⁺	YES
RODGERS (2011)	FUSION PROTEINS	908	UGT CT ⁺ +/-	NO
BUDRYS (2012)	FUSION PROTEINS	908	TFI, FERTILE OR UGT CT ⁺	NO

1.4.4. Serological screen of Gambian trichiasis patients and healthy controls

Using an array of 908 Ct serovar D/UW3-CW proteins produced as GST-fusion proteins ¹⁶¹ sera from Gambian trichiasis patients and controls from a trachoma endemic community were screened (chapter 5)¹⁶². This study identified ten antigens recognised by over half of the fifty-nine samples tested, more significantly four antigens were preferentially recognised by those with trichiasis and eight by healthy controls (Table 1.3). Of particular interest were CT117 (IncF), CT442 and CT556 as they are *Chlamydia*-specific as well as being reportedly immunogenic and protective ¹⁶². IncF and CT442 are both inclusion membrane proteins and they have putative roles in bridging interactions between RBs and the host cytosol through the RER¹⁶³ and modulating host transcription respectively⁸². The study had a number of limitations.

Purification of the GST-fusion proteins was inadequate which made quantification impossible, limiting repeatability and the ability to compare responses to different antigens on the array. In addition it is unclear how the large GST-fusion (approximately 26 kDa) would have impacted native confirmation of the proteins^{164, 165}, which might explain poor immunogenicity of known chlamydial antigens including MOMP (CT681) and HSP60 (CT110). Aside from these technical issues there were also problems with sample selection and analysis. Only 34 case and 25 controls sera with the highest antibody titres against ocular serovar Ct EBs were screened on the complete array. No attempt was made to normalise across the 908 antigens and there was no correction for multiplicity of testing, meaning false positives were not properly accounted for¹⁶⁶.

The overlap of antigen recognition from this study with results from the arrays screening serum and cells from persons with urogenital infection is shown in Table 1.3,

demonstrating a considerable disparity between the findings. This lack of common antigen identification both between and within biovars implies differences may not be entirely due to the different sites of infection, this is unsurprising considering the limited variation between serovars. A more probable cause of the differences is variation in the form of arrays used. Several publications used the same GST-fusion array^{161, 162, 167-169}, one used a similar approach but with histidine-tagged proteins¹⁷⁰, another used affinity-purified proteins printed on slides¹⁷¹, while another used a genomic expression library¹⁷². Even within these approaches the quality of protein produced varied significantly, as did the number of proteins tested, ranging from 100 to 908. Additionally, the criteria for considering an antigen ‘immunodominant’ were not consistent. These discrepancies reiterate the need for validation of the findings from screening the Gambian trachomatous sera.

Table 1.3. Summary of frequently and differentially recognised antigens in adults with TT.

Nine-hundred and eight ORFs were screened with sera from Gambian adults with and without scarring trachoma. ID was based on Ct D/UW3 nomenclature. Antigens recognised by more than half of the sera (50 % recognition), differential clinical association with scarring (pathology) or lack of scarring (protection) and previous identification of each antigen by B-cells and T-cells are indicated.

ID	NAME	>50% RECOGNISED	CLINICAL ASSOCIATION	B-CELL	T-CELL
CT414	PmpC	No	Pathology	1/5 ^{167-169, 171, 172}	0/2 ^{171, 172}
CT667	Hypothetical	No	Pathology	0/4 ^{167-169, 172}	0/1 ¹⁷²
CT695	Hypothetical	No	Pathology	2/4 ^{167-169, 172}	0/1 ¹⁷²
CT706	ClpP2	No	Pathology	0/5 ^{161, 167-169, 172}	0/1 ¹⁷²
CT019	IleS	No	Protection	0/4 ^{167-169, 172}	0/1 ¹⁷²
CT117	IncF	No	Protection	0/5 ^{161, 167-169, 172}	0/1 ¹⁷²
CT301	PknD	No	Protection	0/4 ^{167-169, 172}	0/1 ¹⁷²
CT442	CrpA	No	Protection	3/5 ^{161, 167-169, 172}	0/1 ¹⁷²

				172	
CT553	Fmu	No	Protection	0/4 ^{167-169, 172}	0/1 ¹⁷²
CT556	Hypothetical	No	Protection	0/4 ^{167-169, 172}	0/1 ¹⁷²
CT571	GspE	No	Protection	0/5 ^{161, 167-169, 172}	0/1 ¹⁷²
CT709	MreB	No	Protection	0/4 ^{167-169, 172}	0/1 ¹⁷²
CT089	CopN	Yes	None	6/7 ^{161, 167-172}	0/2 ^{170, 172}
CT143	Hypothetical	Yes	None	2/4 ^{167-169, 172}	0/1 ¹⁷²
CT240	RecR	Yes	None	1/4 ^{167-169, 172}	0/1 ¹⁷²
CT456	Tarp	Yes	None	3/5 ^{167-169, 171, 172}	0/2 ^{171, 172}
CT517	RplX	Yes	None	0/4 ^{167-169, 172}	0/1 ¹⁷²
CT694	Hypothetical	Yes	None	3/5 ^{161, 167-169, 172}	1/1 ¹⁷²
CT806	Ptr	Yes	None	2/5 ^{167-169, 171, 172}	0/1 ¹⁷²
CT841	FtsH	Yes	None	1/5 ^{167-169, 171, 172}	0/2 ^{171, 172}
CT858	CPAF	Yes	None	4/5 ^{161, 167-169, 172}	1/1 ¹⁷²
PCT03	Pgp3	Yes	None	2/4 ^{167, 168, 171, 172}	0/2 ^{171, 172}

1.5. Vaccine trials and current status

Vaccine trials for trachoma began in the 1960's using live-attenuated bacteria isolated from endemic populations, with varied results. Generally they afforded some protection, often short-lived¹⁷³, and reduced disease severity^{174, 175}, importantly showing the potential for a vaccine to protect from trachoma. However, results in non-human primate models and the contemporary interpretation of clinical disease in the human trials suggested that in some circumstances a delayed type hyper sensitivity reaction may have been induced resulting in more severe disease and worse conjunctival

scarring. As a result there have been no new trials in humans since the last trial in India in 1979¹⁷⁶.

1.5.1. MOMP as the primary vaccine candidate

MOMP has been the primary focus of vaccine trials as it induces infection-neutralising antibodies *in vitro*^{132, 177}. Vaccination with various formulations of complete MOMP¹⁷⁸⁻¹⁸² or peptide epitopes¹⁸³⁻¹⁸⁵ induce strong immune responses, however while infectivity has been reduced in some studies, protection has been short-lived and ineffective against the major pathologies^{181, 185}. A need for native conformation¹⁸⁶, its limited protection and recent suggestions that immune responses to MOMP may be host-detrimental³⁴ mean new vaccine targets are needed.

1.5.2. New targets and delivery systems

Recent studies have utilised new delivery systems and new immune targets identified from micro-arrays for B and T-cell antigens, these have also been combined with strategies utilising targeted immunogenic regions of MOMP. Trialled adjuvants and delivery systems include; hepatitis B core antigen, *E. coli* bacterial ghosts¹⁸⁷, cationic liposomes¹⁸⁸, nanoparticles¹⁸⁹ and classical adjuvants such as CpG DNA¹⁹⁰. Members of the polymorphic membrane protein family are immunogenic and partially protective^{187, 191-193}, as are a number of type-three secretion substrates including Tarp, CPAF and CT442 an inclusion membrane protein^{194, 195}. A number of these studies have also employed new formulations of MOMP, including a stabilised trimer supposed to represent native structure¹⁹⁶ and a fusion construct consisting of neutralising epitopes from four different urogenital serovars^{188, 191}.

These new targets and screening methods, combined with the ability to direct vaccines to particular cell types and tissues¹⁹⁷ suggest more effective formulations are possible for vaccination against Ct. They have also highlighted the importance of route of delivery and type of immune response induced. Despite the mixed evidence for protection afforded by antibodies *in vivo*, neutralising antibodies in combination with Th1- type T-cells appear to be necessary for protection.

1.5.3. Current status

There is current optimism as three different vaccines, formulated and delivered in different ways, are approaching phase 1 clinical trials. A combination of neutralising MOMP epitopes from the three most prevalent sexually transmitted serotypes with/without a recombinant MOMP protein delivered by cationic liposomes^{191, 198}, UV-inactivated Ct delivered mucosally conjugated to adjuvant nanoparticles^{189, 199} and plasmid-cured Ct delivered direct to the surface of the eye^{200, 201}. Despite some success in animal models, it is impossible to know whether these vaccines will be effective in humans undergoing natural, possibly repeated, rounds of infection and disease.

2. Hypotheses and Research Questions

2.1. Humoral immunity plays a significant role in protection from human ocular *Chlamydia trachomatis* infection and trachomatous disease

2.1.1. Do serological anti-*Chlamydia trachomatis* responses correlate with resolution of infection and protection from reinfection?

2.1.2. Do serological responses associate with the blinding sequelae of infection in trachoma?

2.1.3. Do antibodies limit progression of scarring in endemic communities?

2.2. Targets of humoral immunity in trachoma are under natural selection

2.2.1. Can we identify new antigenic targets of Ct immunity in trachoma using protein microarrays?

2.2.2. Are the identified antibody targets under immune-driven selection?

2.2.3. Are signatures of selection detectable within a population of ocular Ct strains?

2.3. CT442 is a host immune target in trachoma that is involved in bacterial pathogenesis

2.3.1. What is the intracellular localisation of CT442 and does this facilitate intracellular development and survival of Ct through specific host interactions?

2.3.2. What host and/or chlamydial proteins and pathways does CT442 engage?

3. Methods

3.1. Protein expression and purification

3.1.1. Agarose and SDS-polyacrylamide gel electrophoresis

PCR products and validation steps during cloning of *Chlamydia trachomatis* (Ct)-constructs were visualised by agarose gel electrophoresis. Gene fragments were stained using SYBR Safe (Thermo Fisher Scientific) and band sizes were estimated using 1 Kb DNA ladder (Thermo Fisher Scientific).

Protein samples diluted 1:1 in Laemmli buffer and incubated at 95°C for 10 minutes, were visualised by sodium dodecyl sulphate (SDS)-polyacrylamide gel electrophoresis using either NuSep pre-cast protein gels (Homebush, NSW) or in-house gels for 1.5 to 3 hours using 120 to 140 volts. Bands were visualised using Coomassie blue dye or silver stain, when the protein concentration was expected to be < 0.2 mg.ml. Band sizes were estimated using either Precision Plus Protein Dual Colour (Bio-Rad) or Amersham ECL Full-Range Rainbow (GE Healthcare Life Sciences) molecular weight marker.

3.1.2. Immunoblotting

Samples run on SDS-polyacrylamide gels were transferred to nitrocellulose membranes at 4 °C for 1-1.5 hours using 75 volts. Membranes were washed in washing buffer B (buffer details Appendix 9.1) at room temperature for 30 minutes and blocked three times in blocking buffer D at room temperature for 30 minutes. Membranes were incubated at 4 °C overnight with relevant primary antibody (Table 3.1). After three washes, membranes were incubated at room temperature for 2 hours with either goat anti-mouse (GAM) or goat anti-rabbit (GAR) horse radish peroxidase (HRP) conjugate diluted 1:1000 in blocking buffer D and incubated at room temperature for 2 hours. After a final three washes, membranes were incubated at room temperature for 3 minutes with Amersham ECL Western Blotting Detection Reagent (GE Healthcare Life Sciences). After exposure for 1 to 30 minutes, membranes were developed. Band sizes were estimated as above.

Table 3.1. Primary antibodies for immunoblotting.

PRIMARY ANTIBODY TARGET	DILUTION	HOST SPECIES
CT442	1:1000	Rabbit
GFP	1:5000	Mouse
GST	1:10000	Mouse
HISTIDINE	1:1000	Rabbit
RAB7A	1:1000	Mouse

3.1.3. Provided Ct constructs

27 GST-fusion Ct constructs were provided by Professor Guangming Zhong (UT Health Science Center, San Antonio, TX) and were produced as previously described¹⁶¹.

CT089 and CT875 were provided by Professor Steven Reed (Infectious Disease Research Institute, Seattle, WA) and were produced and expressed as previously described¹⁷².

Both N and C-terminal GFP-fusion CT442 constructs were provided by Dr Maud Dumoux (Birkbeck, UCL, London), these were used for in vitro experiments described in chapter 3.7. Chaperone plasmids pKJE7 and pTf16 were provided by Dr Gerald Larrouy-Maumus (Imperial College, London) and produced commercially by Takara Bio (Kusatsu, Japan), these express different types of molecular chaperones to improve folding of proteins being expressed in *E. coli* (Table 3.2)^{202, 203}.

Table 3.2. Molecular chaperones expressed by the Takara Chaperone Plasmids.

PLASMID	CHAPERONE(S)	RESISTANCE MARKER
PKE7	dnaK, dnaJ, grpE	Chloramphenicol
PTF16	Tig	Chloramphenicol

3.1.4. Transformation and miniprep of provided constructs

For transformation, heat-shock competent or electro-competent *E. coli* and plasmid were defrosted on ice. Two μ l plasmid DNA (pDNA) was added to 50 μ l *E. coli* and incubated on ice for 30 minutes for heat-shock or 2 minutes for electroporation. Bacteria were incubated at 42°C for 2 minutes or electroporated, followed by addition of 950 μ l LB and incubation at 37°C for 1 hour at 230 RPM. Bacteria were pelleted by centrifugation for 1 minute at 13000 RPM, resuspended in 50 μ l LB and spread on agar plates with 50 μ g/ml ampicillin and other required antibiotic. Plates were incubated at 37°C overnight and stored long-term at four °C.

Colonies from transformed bacteria were used to inoculate 100 ml LB with 50 μ g/ml ampicillin and other required antibiotic, cultures were incubated at 37°C for 3 hours at 230 RPM. One ml culture was stored at -80°C with 10% glycerol for future inoculations, the remaining culture was pelleted by centrifugation at four °C for 30 minutes at 4500 RPM. pDNA was purified from pellets using NucleoBond Xtra Midi as described in the manufacturers protocol (Macherey-Nagel). Briefly pellets were resuspended in RES-EF buffer, mixed 1:1 with LYS-EF buffer and incubated on ice for 5 minutes to lyse the bacteria. The lysate was mixed 2:1 with NEU-EF buffer and incubated on ice for 5 minutes followed by application to an equilibrated column. The column was washed once each with FIL-EF, ENDO-EF and WASH-EF buffers. ELU-EF buffer was added to the column, the eluate was mixed two to one with minus 20 °C isopropanol and incubated overnight at 4 °C to precipitate pDNA. pDNA was pelleted by centrifugation at 4 °C for 30 minutes at 13000 RPM and the pellet dried briefly at room temperature. The pellet was washed with 70% ethanol and centrifuged as before except at room temperature. After drying, the pellet was resuspended in an appropriate volume of 50°C H₂O-EF buffer and pDNA concentration was determined using a NanoDrop.

3.1.5. Design and production of constructs

A polyhistidine (His)-tagged CT442 construct was designed using the pET22b(+) expression vector, to produce the protein with a smaller tag that would better reflect the native tertiary structure. Primers for CT442 were designed using Primer3Plus

(<http://primer3plus.com/>), with N-terminal BamHI and C-terminal XhoI restriction enzyme sites (Appendix Table 9.2). CT442 was amplified from genomic Ct DNA by polymerase chain reaction (PCR) using PCRBIO HiFi polymerase (PCR Biosystems), the product was run on an agarose gel and purified using QIAquick Gel Extraction Kit (Qiagen). The ends of the CT442 fragment for cloning and pET22b(+) vector were digested using BamHI and XhoI with NEB buffer (Table 3.3). The digestion products were ligated using T4 ligase and T4 ligation buffer (NEB) (Table 3.4). Ligated plasmid was transformed into DH5 α *E. coli* as described above. Colonies were screened for successful transformation by colony PCR.

Table 3.3. Trialled conditions for restriction digest of the CT442-His insert.

DIGESTION TYPE	BUFFER	INCUBATION TIME (HOURS)	VECTOR DEPHOSPHORYLATION
DOUBLE	NEB 3.1	1	No
DOUBLE	NEB CutSmart	1	No
DOUBLE	NEB CutSmart	1	Yes
DOUBLE	NEB 3.1	2	No
DOUBLE	NEB 3.1	12	No
DOUBLE	NEB 3.1	1	No
DOUBLE	NEB CutSmart	2	Yes
SEQUENTIAL	NEB CutSmart	1	Yes
SEQUENTIAL	NEB CutSmart	2	Yes

Table 3.4. Trialled conditions for ligation of the CT442 insert into the pET22b(+) plasmid.

INSERT VECTOR RATIO	TO INCUBATION TIME (HOURS)	INCUBATION TEMPERATURE (°C)	TRANSFORMATION METHOD
3:1	0.5	37	Heat-shock
3:1	0.5	20	Heat-shock
3:1	3	20	Heat-shock
3:1	12	4	Heat-shock
3:1	0.5	37	Heat-shock
3:1	3	20	Heat-shock
3:1	12	4	Heat-shock
6:1	0.5	37	Heat-shock
6:1	12	4	Heat-shock
3:1	12	4	Electroporation
6:1	0.5	37	Electroporation
6:1	12	4	Electroporation

Due to difficulties in ligating the CT442 gene into pET22b(+) expression vector, the construct was produced commercially through GenScript (Piscataway, NJ).

3.1.6. Optimisation of expression and purification conditions

A small volume of LB media (2x YT for GST-constructs) was inoculated with frozen stocks of transformed *E. coli* and incubated at 37°C overnight at 230 RPM. Overnight cultures were diluted 1:50 into applicable media and incubated at 37°C at 230 RPM until OD 600 reached 0.6 to 1.2. IPTG was added at a concentration from 100 µM to 1 mM to induce expression and incubated at 10°C to 37°C from 2 hours to overnight.

Cultures were centrifuged at 4 °C for 30 minutes at 4500 RPM, the pellets were resuspended in 1/20th culture volume of Tris-buffered saline (TBS, PBS for GST-constructs) with protease inhibitors. The bacteria were lysed by cell disruption followed

by centrifugation at 4 °C for 30 minutes at 4000 RPM to remove aggregated material and cell debris. Relevant detergents were added and the samples were incubated at room temperature for 2 hours on a roller. The samples were centrifuged at 4 °C for 1 hour at 160000 RPM, the supernatants contained the soluble proteins and were stored at -20°C. The pellets contained insoluble proteins and was either discarded or solubilisation was attempted using different detergents (Table 3.5).

Table 3.5. Trialled detergents for improved solubility of expressed proteins.

NAME	DETERGENT TYPE
ASB-14	Zwitterionic
CHAPS	Zwitterionic
DDM	Non-ionic
LDAO	Zwitterionic
OCTYL B-D-1-THIOGLUCOPYRANOSIDE	Non-ionic
POLYOXYETHYLENE (10) TRIDECYL ETHER	Non-ionic
TRITON X-100	Non-ionic

To purify GST-fusion constructs 1/2000th culture volume of 50% glutathione sepharose beads were added to the soluble protein samples to bind the GST moiety and incubated at room temperature for 2 hours on a roller. The samples were passed down a 5 ml polypropylene column leaving a bed of GST-fusion bound beads. The beads were washed five times with 5 ml PBS before adding 1/4000th culture volume elution buffer and incubated at room temperature for 30 minutes. The eluates were collected and the elution step repeated twice.

To cleave the GST-fusion prior to purification the process was repeated until the beads were passed down the column. The beads were washed five times with 5 ml cleavage buffer before adding 1/4000th culture volume cleavage buffer with 8% PreScission Protease (GE Healthscare Life Sciences) and incubated at four °C overnight. The eluates were collected and the elution step repeated twice.

GST-fusion proteins and cleaved proteins were further purified by size-exclusion based gel-filtration chromatography using a 35 ml Superdex 200 column (GE

Healthcare Life Sciences) with an ÄKTApurifier. The column was washed with 1 column volume distilled H₂O followed by one column volume PBS containing detergents found in the samples being purified. The samples were applied to the column and 1 further column volume PBS with detergent was applied to the column with fractions taken every 0.5 ml.

His-tagged CT442 was purified by affinity chromatography using a 5 ml HisTrap (GE Healthcare Life Sciences) with an ÄKTApurifier. The column was washed with one column volume distilled H₂O followed by 1 column volume TBS containing detergents found in the samples being purified and 10 mM imidazole. The samples were applied to the column followed by 2 column volumes TBS with step-wise increasing of imidazole concentration up to 250 mM. Weakly and non-specifically bound proteins should be washed from the column with low concentrations imidazole, proteins with a high affinity should remain bound until the imidazole concentration increases. Fractions were taken every 0.5 ml during the imidazole gradient.

3.1.7. Peptide selection and production

Biotinylated peptides of CT442 (AA 135-150), IncA (CT119, AA 258-273) and IncG (CT118, AA 149-164) were produced by thinkpeptides (ProImmune, Oxford, UK), using sequences from Ct D/UW3. Purity was determined by High Performance Liquid Chromatography (HPLC), the minimum accepted was 80 %. Lyophilised peptides were resuspended in distilled H₂O, then aliquoted and stored at -80°C. All peptide sequences had been previously identified as immunogenic (personal communications, Dr Bernhard Kaltenboeck)²⁰⁴.

3.2. Serological screening of proteome arrays

3.2.1. Glutathione S-transferase fusion protein micro-titre plate array

A glutathione S-transferase (GST) fusion protein enzyme-linked immunosorbent assay (ELISA) was used to screen the Ct proteome with sera from 59 people (chapter 5). The original results were published by Lu et al¹⁶², the details of the array were published by

Sharma et al¹⁶¹. Nine-hundred and eighteen open reading frames (ORFs) from Ct serovar D/UW3-CW, encompassing the genome and plasmid, were cloned into the pGEX expression vector which produces protein with an N-terminal GST-fusion (Amersham Biosciences Corp., Piscataway, NJ), these were a combination of full-length and N/C-terminal fragments (each half full-length). After transformation into *E. coli* expression of each protein was optimised for isopropyl- β -D-1-thiogalactopyranoside (IPTG) concentration (0.1 to 5 mM), bacterial density at induction (0.5 to 1.5 optical density [OD]), incubation temperature (10 °C to 30 °C) and time (0.5 hours to overnight). After centrifugation the bacteria pellets were resuspended into Triton lysis buffer and lysed by sonication, debris was removed by high-speed centrifugation and the supernatant was frozen at -80 °C. Successful expression of the proteins was determined by purification on glutathione-conjugated agarose beads and running on sodium dodecyl sulphate (SDS)-polyacrylamide gels. Lysates were accepted for the micro-titre plate array if a band was present at the expected weight after staining with Coomassie blue dye. In total 933 lysates were used, covering 908 serovar D Ct proteins.

Bacterial lysates were diluted 1:10 in PBS, added 200 μ l/well into glutathione-coated 96-well microplates (Pierce, Rockford, IL) and incubated overnight at four °C to allow binding. Lysate from an empty pGEX vector (containing free GST) and GST-chlamydial protease/proteome-like activity factor (CPAF) were included as positive and negative controls respectively. To reduce background absorbance from non-specific antibodies, human sera to be tested were incubated overnight at 4 °C with lysate from an empty pGEX vector followed by purification on glutathione-conjugated agarose beads. Prior to testing, the plates were washed twice with washing buffer A and blocked with blocking buffer A for 1 hour at room temperature. Serum samples were diluted in blocking buffer B and added to the plates for 2 hours at room temperature. After washing, human antibody reactivity was detected by adding of alkaline phosphatase-conjugated goat anti-human immunoglobulin G (IgG; Jackson ImmunoResearch Laboratories, Inc., West Grove, PA) and the substrate *p*-nitrophenyl phosphate. The antibody binding was measured as absorbance (OD) at 405 nm.

For determining positive antibody responses all sera were separately incubated overnight at 4 °C with lysate from HeLa cells and Ct-infected HeLa cells. Only antibody responses which were unaffected by incubation with HeLa cells but were

significantly lowered by incubation with Ct-infected HeLa cells were considered positive.

3.2.2. Microarray chips

Ct protein microarray chips were prepared as described previously²⁰⁵ by Antigen Discovery (Irvine, CA) and screened using sera from 123 patients (chapter 4). Briefly 894 ORFs from Ct serovar D/UW3-CW genome were PCR amplified and *in vivo* cloned into the pXT7 expression vector which produces proteins with an N-terminal His fragment and a C-terminal haemagglutinin sequence and T7 terminator. Ct-specific products were expressed from the plasmids using an *in vitro* transcription translation system (RTS 100 kit, Roche) and printed on nitrocellulose coated glass slides (GraceBio) using an OmniGrid Accent microarray printer (Digilab). Successful expression of the proteins was determined using antibodies against His (clone His-1; Sigma) and haemagglutinin (clone 3F10; Roche). Eight-hundred and sixty-four of 894 ORFs were successfully expressed, determined by positivity for both terminal tags²⁰⁵.

Prior to testing sera was diluted 1/100 in blocking buffer C at room temperature for 30 minutes with agitation while the microarrays were rehydrated using protein array blocking buffer (Whatman, Piscataway, NJ). The arrays were probed with sera at room temperature for 2 hours with agitation. After 3 washes with washing A buffer the microarrays were incubated with biotin-conjugated goat anti-human antibody (Jackson ImmunoResearch Laboratories, West Grove, PA). After 3 further washes the microarrays were incubated with streptavidin-conjugated Sensilight P3 (Columbia Biosciences, Columbia, MD). The microarrays were scanned in a ScanArray Express HT microarray scanner (Perkin Elmer, Waltham, MA) and the fluorescence signal was quantified and corrected for background noise using QuantArray software (Perkin Elmer, Waltham, MA).

3.2.3. Normalisation and filtering

Distribution of the data was determined using the mean of each serum across all tested antigens in a Shapiro-Wilk test and by visualisation of the data in histograms. Normalisation methods were assessed using a published method of ranking to select the

method which best minimised deviation across the array²⁰⁶. To calculate relative rank deviation (RRD), sera were ranked based on mean signal intensity and the absolute standard deviation at each rank was divided by the mean at each rank. In this context each rank consists of one individual.

Subsequent normalisation steps were; selection of included serum, selection of included control antigens, inverse hyperbolic sine transformation and comparison of mean-centring, 2% trimmed mean and 10% trimmed mean. At each step the RRDs were visualised to identify the best method, which produced the lowest RRD. To determine the degree of normalisation achieved, the data was tested as before in a Shapiro-Wilk test and by visualisation.

Post-normalisation the global median of each array was calculated, individual antigens whose median was lower than the global median were excluded. This filtering was done to exclude infrequently and weakly recognised antigens, as they are less likely to be repeatable with different methods and populations.

3.2.4. Determining positivity

In the published analysis of the GST-fusion array positivity was defined as an optical density (OD) equal to or greater than two standard deviations above the mean from the relevant 96-well plate¹⁶². There is no consensus method for what constitutes a significant positive in serological microarrays. There are always samples clearly positive or negative, towards the outer limits of OD/signal intensity, but there is rarely a clear separation of responses into distinct populations as is favoured for clinical assays. Several different methods were tested for identifying breakpoints in the data. The intention was to find the most appropriate method for separating responses to each antigen.

The tested methods included extrinsic and intrinsic methods. Extrinsic methods tested were; mean of no DNA controls, global mean and global mean plus 2 standard deviations. Intrinsic methods tested were; k-means clustering, k-medoids clustering, fuzzy c-means clustering, hierarchical clustering and mixture modelling. The intrinsic methods were tested allowing for 2 to 10 clusters. The average silhouette width of each

antigen was used to determine how well the data fit within its cluster. Silhouette ranges from minus one to one and is defined by equation 1.

$$\text{silhouette}(i) = (b(i) - a(i)) / \max\{a(i), b(i)\} \quad (1)$$

Where i is a data point, a is average dissimilarity with all other data points in its cluster, b is the lowest average dissimilarity to any other cluster of which i is not a member. If i is similar to other data in its cluster a will be low. If i is also dissimilar to data in the nearest cluster b will be high. In this case silhouette will tend towards one. If i is not similar to its cluster a will be high. If i is similar to data in the nearest cluster b will be low. In this case silhouette will tend towards minus one. A silhouette close to one suggests that i is on the edge of two clusters (Table 3.6).

Table 3.6. Summary of how similarities within and between clusters impact the silhouette score.

OWN CLUSTER FIT	A	NEAREST CLUSTER FIT	B	CONCLUSION	SILHOUETTE
GOOD	Low	Bad	High	Fits cluster	Close to 1
BAD	High	Good	Low	Fits nearest cluster	Close to -1
GOOD	Low	Good	Low	On the border	Close to 0
BAD	High	Bad	High	On the border	Close to 0

The mean of each silhouette per antigen gave the average silhouette width, which showed how tightly clustered the data was as a whole. To determine positive responses, two clusters were identified and method which had the highest average silhouette width for each antigen separately was identified. Data points clustered with the maximum OD/signal intensity point of each antigen were considered positive and the opposing cluster negative.

3.2.5. Choice of outcome variable and covariates

Where possible the OD/signal intensity values were used for analysis rather than data categorised into positive and negative responses. This was used to avoid unnecessarily losing information and because most responses were continuous.

For the GST-fusion micro-titre plate array all comparisons were between scarring cases (trachomatous scarring [TS], trachomatous trichiasis [TT] or corneal opacity [CO]) and healthy-matched controls unless otherwise stated. In the published analysis intensity of responses in scarring cases and healthy controls were compared using Student's *t*-test and number of positive responses using Fisher's exact test¹⁶². In the re-analysis intensity of responses were compared using a general linearised model (glm) and number of positive responses using χ^2 test. For the glm 10,000 permutations of the outcome variable were performed to get an adjusted p-value. For cross-validations, summary p-values were calculated using Fisher's method. Likelihood ratio tests were used to compare models with null models only including the covariates, age and gender. Multivariate regression was performed using random forests through the R package ranger, variable importance was determined using the Gini impurity index. Adjusted odds ratios (OR*) and adjusted confidence intervals (CI*) were calculated as the unadjusted OR exponentiated by half the range for a given antigen.

ROC curves were produced manually from specificities and sensitivities computed using R packages pROC and ROCR. ROC curves and AUC classified as random were calculated by 100 random resampling of variables included in the model to obtain a normal distribution from which to draw a median value.

For the microarray chips an episode of infection was defined as a Ct-positive result from an in-house 16s qPCR assay²⁰⁷, an episode of disease was defined as presence of either follicular trachoma (TF) or inflammatory trachoma (TI)²⁰⁸. An episode was considered continuous if an individual's infection or disease status was consistent in consecutive visits. Where data was missing between visits with consistent infection or disease status it was considered continuous, where data was missing between visits with inconsistent infection or disease status it was assumed status

changed at the midpoint. To best utilise the longitudinal information a number of categorisations were examined as potential outcome variable (Table 3.7).

Table 3.7. Different outcome measures considered for analysis of the longitudinal study of Ct infection.

OUTCOME VARIABLE	NUMBER OF CATEGORIES	CATEGORISATION PARAMETERS			NUMBERS PER CATEGORY		
		0	1	2	0	1	2
INFECTIOUS EPISODE(S) OR NOT	Two	No episodes	≥ 1 episode	NA	23	67	NA
NUMBER OF INFECTIOUS EPISODES	Two	No episodes	1 episode	≥ 1 episode	23	19	48
MEDIAN DURATION OF INFECTIOUS EPISODES	Three	No episodes	$0 \leq 2$ weeks	> 2 weeks	23	37	30
MEDIAN DURATION OF INFECTIOUS EPISODES	Two	≤ 2 weeks	> 2 weeks	NA	60	30	NA

Statistical analyses comparing intensity of responses and number of positive responses between the dichotomous outcomes were performed as described previously for the GST-fusion array. For both arrays the influence of covariates was modelled through univariate-glm. Covariates with a significant association with the outcome variable (p-value ≤ 0.1) were included in the glm of the individual antigens. Age and gender were always included.

3.2.6. Diversity metrics

Ecological measures of diversity rely on species breadth/richness, the total number of species in a sample, and species diversity, which additionally incorporates the relative abundance of each species. In this analysis antigens were considered as species, abundance as the response to each antigen and the samples were either the continuous data or split into the dichotomous outcome variables. These definitions are based on the assumption that responses on the array broadly correlate with abundance of circulating antibodies in each sample. A normalised OD of one was interpreted as one arbitrary unit of circulating antibody. This means if a response to an antigen is twice as high in one sample compared with another, circulating antibodies are twice as abundant in that individual.

Breadth was defined as the number of antigens each individual had a positive response to. For the remaining measures examining diversity existing methods were adapted to incorporate the continuous OD/signal intensity values. This was deemed more appropriate as an assumption of these methods was that individuals within a species are equivalent^{209, 210}, in this analysis the species are antigens and positive responses within them are not equal.

Three different measures of diversity were utilised to improve reliability of the results. Shannon's entropy (H) defined by equation 2, Simpson's index (D) defined by equation 3 and Hill numbers defined by equation 4²¹⁰. Higher values for all three indicate increased diversity and greater evenness. High values of H mean that an unknown individual could belong to any species, in our context this means one unit of antibody in a sample could be targeted against any antigen because responses in the sample are even²¹¹. High values of D mean that two randomly chosen individuals are likely to be from different species, in our context this means that if we take two separate unit of antibody from the array they are unlikely to be targeted against the same antigen due to evenness of the responses²¹¹.

$$H = - \sum_{i=1}^s p_i \log p_i \quad (2)$$

$$D = 1 - \sum_{i=1}^s p_i^2 \quad (3)$$

Where S is number of antigens and p_i is the proportion of antibodies specific to each antigen. P_i is estimated as the amount of antibody specific to each antigen divided by the total amount of antibody present in each individual.

Hill numbers are slightly more complex because they change with order of diversity (q), as q increases more weight is placed on the most abundant species or in our case the most strongly recognised antigens²¹⁰. Within each value of q , higher values indicate increased diversity. Of greater interest is the diversity profile, which is how Hill numbers change with increasing q . The steeper the decline the more uneven the sample is, in our context a steeper decline in Hill numbers implies antibody responses are focussed on a few strongly recognised antigens.

$$\text{Hill number } (q) = \left(\sum_{i=1}^S p_i^q \right)^{1/(1-q)} \quad (4)$$

Where S is number of antigens, p_i is the proportion of antibodies specific to each antigen and q is the order of diversity. Figure 3.1 provides a simplified interpretation of breadth, diversity and evenness, A) represents an even sample where responses are not focussed on one antigen and B) represents an uneven sample where diversity is less because responses are predominantly targeted against one antigen.

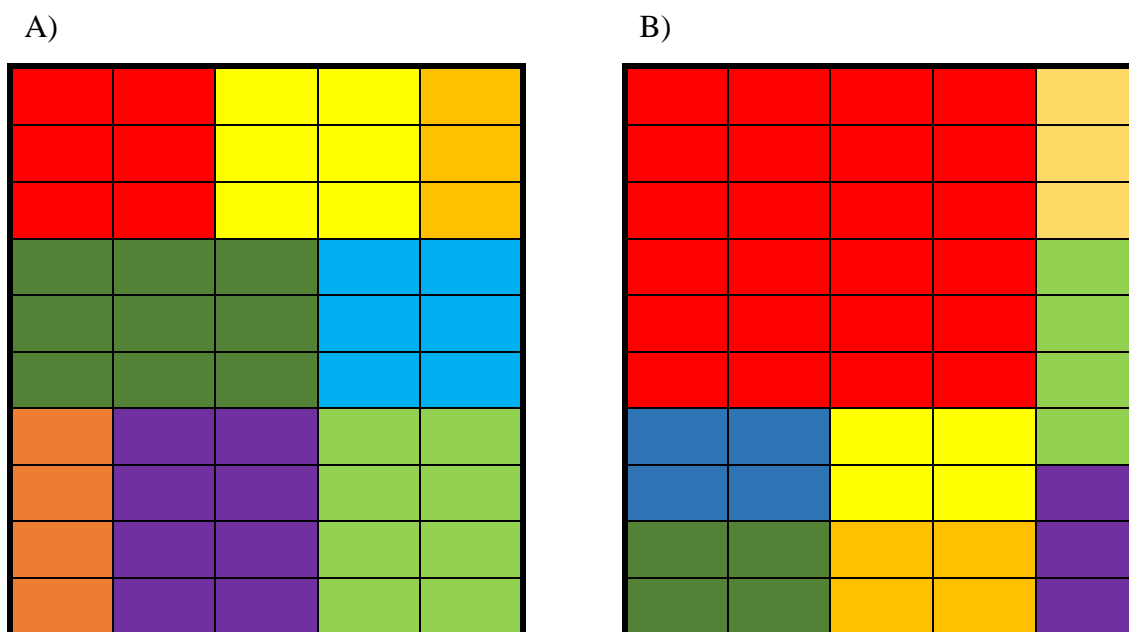


Figure 3.1. Visualisation of theoretical antibody responses in two samples, different colours represent abundance of antibody responses to different antigens.

A) High levels of breadth and diversity. B) Similar breadth however diversity is reduced. In both samples antibodies against eight antigens are detectable, however in the right-hand sample the majority of antibodies are targeted against one antigen (red). This is an example of an uneven antibody response.

3.3. *In silico* analyses

3.3.1. Expression, localisation and structure

Developmental cycle expression stage for each transcript was based on data and groupings from Belland et al²¹², this grouping was manually assigned to data from Nicholson et al²¹³. Localisation of expressed proteins was predicted using Cello²¹⁴, pSORTB²¹⁵ and loctree3²¹⁶, three of the top performing servers for bacterial proteins²¹⁶. Predicted localisation was defined as the consensus from the 3 predictions. Transmembrane domains were assigned from consensus predictions from MemBrain²¹⁷, Phobius (<http://www.ebi.ac.uk/Tools/pfa/phobius/>) and TMHMM²¹⁸, which have been independently validated²¹⁹.

3.3.2. B and T-cell epitopes predictions

MHC class-1 epitopes were predicted with SYFPEITHI (<http://www.syfpeithi.de/bin>), Net-MHC1²²⁰ and IEDB-SMM²²¹. MHC class-2 epitopes were predicted with SYFPEITHI, Net-MHC2²²² and ProPred. B-cell epitopes were predicted using BepiPred²²³, ABCpred²²⁴ and IEDB-antibody epitope prediction. Consensus predictions from the first 2 servers were matched with regions predicted to contain antibody epitopes based on a consensus of; beta-turn propensity²²⁵, accessibility²²⁶, flexibility, antigenicity²²⁷ and hydrophilicity²²⁸. These have been previously validated as top performers for linear B-cell epitope prediction^{223, 229, 230}.

3.4. Study details of tested serum samples

3.4.1. HLA polymorphism and scarring trachoma

In The Gambia in 1995, 153 people with evidence of TS by clinical examination were recruited alongside age, sex and village-matched controls with normal eyes from Kaur Health Centre and the villages of Jali and Berending. A 1 ml sample of venous blood was taken from each person to obtain serum. Details of the collection are available from Conway et al^{231, 232}. A total of 231 serum samples were tested from this study (Table 3.8). The joint Scientific and Ethics Committee of the Gambian government and the Medical Research Council Laboratories approved the study.

Table 3.8. Sample characteristics from the HLA polymorphism and scarring trachoma study.

	HEALTHY CONTROLS	SCARRED CASES	ASSOCIATED P-VALUE
NUMBER	116	115	NA
AGE IN YEARS	37.50	38.00	0.900
(95% CI)	(7.00-65.00)	(7.00-65.75)	
FEMALE (N [%])	84 (72.41)	80 (69.57)	0.633

3.4.2. NK-cells and trachoma

As part of a study looking at NK-cell responses in West Kiang District of The Gambia in 2006, twelve to fifteen healthy participants from three different age groups (two to five; fifteen to 25 and >35 years) were recruited from trachoma-endemic communities²³³. A total of 37 serum samples were available of which 32 had been previously examined for presence of Pgp3 antibodies²³⁴. The joint Scientific and Ethics Committee of the Gambian government and the Medical Research Council Laboratories approved the study.

3.4.3. Longitudinal cohort study of Ct infection and active trachoma

In 2002 a rapid assessment survey of adults and children was carried out in the Western and North Bank Regions of The Gambia and villages with greater than 20 % prevalence of active trachoma (TF and/or TI) were selected as study sites^{142, 235}. Initially six villages were selected and a further three were included to increase the power of study as the prevalence of active trachoma in the target population (children aged 4-15 years) was lower than expected. The joint Gambian Government-Medical Research Council Ethics Committee and the Ethics Committee of the London School of Hygiene & Tropical Medicine approved the design and procedures of this study.

School-age children in these villages were examined for the clinical signs of trachoma. A subset of 345 children between the ages of 4 and 15 years were recruited and followed for a period of 28 weeks. At baseline and approximately fortnightly visits (ten to nineteen days), children were examined for signs of active trachoma and a digital photograph of each eyelid was taken. Two swabs were collected, one into a dry polypropylene tube and the other into RNeasyTM. Tear fluid was collected from the right eye using a sponge-tipped eye spear (Merocel®, Xomed Surgical Products, Jacksonville, Florida USA), inserted in the inferior conjunctival fornix and held there for approximately 30 seconds. A sample of venous blood was taken at the beginning and cessation of the study. Details of the collection and loss to follow-up are available from Faal et al¹²⁸. An episode of infection was defined as a positive result from an in-house 16S rRNA PCR²⁰⁷, clinical disease was defined according the WHO simplified

grading system²⁰⁸. Collectively, 130 serum samples from baseline and cessation of the study were available for testing (Table 3.9).

Table 3.9. Sample characteristics from the longitudinal cohort study of Ct infection and active trachoma.

	PROTECTED	UNPROTECTED	ASSOCIATED P-VALUE
NUMBER	95	35	NA
AGE IN YEARS	9.00	9.00	0.982
(95% CI)	(2.00-14.00)	(1.85-13.15)	
FEMALE (N [%])	42 (44.21)	13 (37.14)	0.470
VILLAGE (N [%])			0.002
- 1	13 (13.68)	18 (51.43)	
- 2	10 (10.53)	2 (5.71)	
- 3	1 (1.05)	0 (0.00)	
- 4	14 (14.74)	3 (8.57)	
- 6	14 (14.74)	5 (14.29)	
- 7	34 (35.79)	6 (17.14)	
- 8	9 (9.47)	1 (2.86)	

3.4.4. Scarring case-control study

In The Gambia from May 2006 to February 2009, 61 people with evidence of TS and TT by clinical examination were recruited alongside age, sex and village-matched controls with normal eyes from the Western, Central and Lower River Regions. A sample of venous blood was taken from each person to obtain serum. Details of the collection are available from Lu *et al* ¹⁶². A total of 116 of these archived sera were available for testing (Table 3.10). The study and its procedures were approved by the joint Gambian Government/Medical Research Council Ethics Committee

Table 3.10. Sample characteristics from the 2006-2009 scarring case-control study.

	HEALTHY CONTROLS	SCARRED CASES	ASSOCIATED P-VALUE
NUMBER	58	58	NA
AGE IN YEARS	55.50	60.00	0.199
(95% CI)	(30.43-73.73)	(34.00-77.88)	
FEMALE (N [%])	40 (68.97)	39 (67.24)	0.842

3.4.5. NK-cells and scarring trachoma

In the Gambia in 2011, 90 people with evidence of TS and TT by clinical examination were recruited alongside age, sex and village-matched controls with normal eyes from multiple rural regions⁵⁵. A sample of venous blood was taken from each person to obtain serum. A total of 90 of these serum samples were available for testing. The Ethics Committee of the Gambian Government/Medical Research Council Unit, and the ethics committee of the London School of Hygiene and Tropical Medicine approved the study

Table 3.11. Sample characteristics from the NK-cells and scarring trachoma study.

	HEALTHY CONTROLS	SCARRED CASES	ASSOCIATED P-VALUE
NUMBER	38	52	NA
AGE IN YEARS	19.50	20.50	0.224
(95% CI)	(1.00-39.00)	(3.55-37.73)	
FEMALE (N [%])	30 (78.95)	45 (86.54)	0.343

3.4.6. Longitudinal cohort in Tanzanian children investigating scarring progression (2012-2016)

The study was conducted in three adjacent trachoma endemic communities in Kilimanjaro and Arusha regions, Northern Tanzania. A cohort of children aged 4 to 12

years were followed for 4 years. At baseline (January 2012) and every three months, children were examined for clinical signs of trachoma. Conjunctival swabs were taken to test for Ct infection. Ct DNA was detected using a previously described droplet digital PCR assay²³⁶.

At cessation of the study blood samples were collected from a subset of 372 children. Finger prick blood specimens were collected onto filter paper which hold approximately 10 µL of blood. Filter papers were air-dried on a dedicated specimen rack for between four and twelve hours, before being sealed in a small zip lock bag. Field-grades for scarring at baseline and the end of the study were regraded from digital photographs by an ophthalmologist experienced in trachoma grading, scarring progression was defined as the presence of new conjunctival scars when comparing baseline and end-stage photographs side by side (Table 3.12). Some children were not seen at baseline or at the end of the study, these 61 were excluded from the analysis of scarring progression. Blood-spot ELISA was performed in the Kilimanjaro Clinical Research Institute (KCRI) laboratory, Kilimanjaro Christian Medical Centre (KCMC), Moshi, Tanzania. This study was reviewed and approved by the Tanzanian National Institute for Medical Research Ethics Committee, the Kilimanjaro Christian Medical Centre Ethics Committee, and the London School of Hygiene and Tropical Medicine Ethics Committee.

Table 3.12. Demographics of children sampled at the end of a 4-year longitudinal cohort of scarring progression in Tanzania.

	NO PROGRESSION	PROGRESSION	P-VALUE
NUMBER	243	68	NA
AGE IN YEARS (95% CI)	11.00 (9.00-16.00)	11.00 (9.00-15.33)	0.465
FEMALE (N [%])	139 (57.20)	46 (67.65)	0.123
BASELINE CT⁺ (N [%])	28 (13.53)	12 (20.69)	0.181
VISITS WITH F SCORE > 1 (N [%])			0.027
0	116 (47.74)	24 (35.29)	

1-2	64 (26.34)	15 (22.09)	
> 2	63 (25.92)	29 (45.31)	
VISITS WITH P SCORE > 1 (N [%])			< 0.001
0	165 (67.90)	20 (29.41)	
1-2	55 (22.63)	21 (30.89)	
> 2	23 (9.47)	27 (39.71)	
BASELINE SCARRING GRADE (N [%])			< 0.001
0	171 (70.37)	25 (36.76)	
1	21 (8.64)	15 (22.06)	
2	4 (1.65)	7 (10.29)	
3	11 (4.53)	11 (16.18)	
FINAL SCARRING GRADE (N [%])			< 0.001
0	219 (90.12)	0 (0.00)	
1	12 (4.94)	17 (25.00)	
2	4 (1.65)	13 (19.12)	
3	8 (4.94)	38 (55.88)	

3.4.7. NK-cells and HCMV, cross-sectional survey (The Gambia 2011)

In The Gambia, participants were recruited from the villages of Keneba, Manduar, and Kantong Kunda in the West Kiang district for a study investigating NK-cell function in HCMV²³⁷ infections and the modulating effects of vaccination on their functions. Venous blood samples were collected from individuals aged 1 to 49 years, serum samples from 495 of these individuals were available for testing. This study was approved by the ethical review committees of the Gambia Government/Medical Research Council and the London School of Hygiene and Tropical Medicine.

3.4.8. The Solomon Islands Trachoma prevalence survey 2014

In the Solomon Islands, 13 villages in Temotu, Rennell and Bellona provinces were selected based on a pre-MDA survey in September-November 2013 showing clinical signs of TF in over 10% of the population²³⁸. Finger prick blood spots were collected from 1501 people of all ages as described above. The study was approved by the London School of Hygiene & Tropical Medicine Ethics Committee and the Solomon Islands National Health Research Ethics Committee.

3.5. Humoral immunology

3.5.1. Optimisation and validation of the in-house ELISA

Initial optimisation of this ELISA method was achieved using the positive control antigen Pgp3. Serum dilution, anti-human IgG-peroxidase antibody concentration (Sigma Aldrich) and antigen concentration were optimised. 32 samples from the NK-cells and study (chapter 3.4.2) were previously tested using an independent Pgp3-based serological assay published by Wills *et al*²³⁴, OD values from that assay were used as reference or ‘gold’ standard to test the in-house ELISA. The optimised ELISA was further validated on 495 sera representing individuals with ages 1 to 49 years from a Gambian community endemic for trachoma (chapter 3.4.7).

For testing blood-spot eluates rather than serum/plasma samples, and for additional validation, the in-house ELISA was used to test Pgp3 responses in 260 samples collected from a trachoma prevalence survey in the Solomon Islands (chapter 3.4.8)²³⁸. Results were compared with those from an independent Pgp3-based serological assay developed at the Centre for Disease Control (Atlanta, GA) and currently being employed by the Global Trachoma Mapping Project²³⁹.

3.5.2. Serum/plasma ELISA

Antigens were diluted to one µg/ml in coating buffer A (for proteins) or coating buffer B (for peptides) and 50 µl/well added to the first 90 wells of Immulon 4 HBX microtitre plates (Fisher Scientific, Loughborough, UK). Pgp3 was similarly diluted and added to the remaining 6 wells of each plate as a positive control. Antigens were bound at 4 °C

overnight, the plates were covered throughout the protocol. Sera/plasma were diluted 1/500 (for proteins) or 1/250 (for peptides) in blocking buffer A, samples stored in glycerol were tested twice as concentrated. Four hyper-immune control sera were pooled and diluted 1/100 and then serially diluted 1/5, 5 times. After 30 minutes agitation, diluted samples were stored at 4 °C overnight.

The following day plates were inverted and washed twice with washing buffer A and blocked with 100 µl blocking buffer A at room temperature for 4 hours. After 2 washes 100 µl test sera in triplicate and control sera were added and incubated at room temperature for 4 hours. After 4 washes 100 µl anti-human IgG-peroxidase antibody diluted 1/30000 blocking buffer A were added per well and incubated at room temperature for 1 hour. After a final 4 washes 100 µl 1-Step Ultra TMB-ELISA substrate (Fisher Scientific, Loughborough, UK) was added per well and incubated at room temperature for 10 minutes (for proteins) or 15 minutes (for peptides). The reaction was stopped by addition of 100 µl 2 M sulphuric acid per well to stop the reaction and the plate read at OD 450 nm for detection and 700 nm for background correction.

3.5.3. Dried blood-spot ELISA

Blood-spots were picked and added to individual wells of 96-deep well plates, antibodies were eluted in 300 µl blocking buffer B at 4 °C overnight. Plates and control sera were otherwise prepared as above. The following day plates were washed and blocked as above. Forty µl blood-spot eluates and 100 µl control sera were added per well and incubated at room temperature for 2 hours. The antibody binding and detection were performed as above.

3.5.4. Biotinylated peptide ELISA

Stocks of streptavidin were diluted to 5 µg/ml in H₂O and 100 µl/well added to the first 90 wells and Pgp3 was added to the remaining 6 wells on each plate as above. Streptavidin and Pgp3 were dried onto the plates by incubation, uncovered, at 37°C overnight. Blood-spots were diluted as above, sera/plasma were diluted 1/250.

The following day the plates were rehydrated with washing buffer A at room temperature for 15 minutes. Biotinylated peptides were diluted to 1 µg/ml in coating buffer B and 50 µl/well added to the first 90 wells, blocking buffer A was added to the remaining 6 wells of each plate. Peptides were bound to streptavidin at room temperature for 1 hour with agitation. After 2 washes plates were blocked with 100 µl blocking buffer A at room temperature for 30 minutes. After two washes the antibody binding and detection were performed as above.

3.5.5. Between-plate normalisation

Non-specific absorbance at OD 700 nm was subtracted from absorbance at 450 nm, samples in triplicate were averaged and values greater than 1 standard deviation from the mean were excluded. Values from each positive control dilution were averaged across all plates for each sample set. Values from each plate were divided by the averaged values, the mean deviation for each plate's serial dilution from the average was used to transform each plates values.

3.5.6. Comparison of OD values

Normalised OD values were compared as described previously. Briefly, positivity was determined using an objectively determined clustering method and number of positives compared using χ^2 test. Continuous OD values were compared using a glm.

3.6. Population genomics

3.6.1. Isolate collection and whole-genome sequencing

Survey, clinical examination and sample collection methods have been described previously²⁴⁰. Briefly, we conducted a cross-sectional population-based survey in trachoma-endemic communities on the Bijagós Archipelago of Guinea Bissau. Conjunctival swabs were obtained from the left upper tarsal conjunctiva of each

participant, DNA was extracted and Ct omcB (genomic) copies/swab quantified from the second conjunctival swab using droplet digital PCR (ddPCR)²³⁶.

For 8 individuals, whole genome sequence (WGS) data was obtained following Ct isolation in cell culture. Briefly, samples were isolated in McCoy cell culture by removing 100 µl eluate from the original swab with direct inoculation onto a glass coverslip within a bijoux containing Dulbecco's modified Eagles' Medium (DMEM). The inocula were centrifuged onto cell cultures at 1800 rpm for 30 minutes. Following centrifugation the cell culture supernatant was removed and cycloheximide-containing DMEM added to infected cells which were then incubated at 37°C in 5% CO₂ for three days. Viable Ct elementary bodies (EB) were observed by phase contrast microscopy. Cells were harvested and further passaged every three days until all isolates reached a multiplicity of infection between 50-90 % in 2 x T25 flasks. At this point cells were harvested, scraped and centrifuged to concentrate and remove cell debris. All isolates of Ct were negative when tested for mycoplasma contamination by fluorescence microscopy using Hoechst 33258 staining and by using a VenorGem Mycoplasma PCR Detection Kit (Minerva Biolabs, Berlin, Germany) according to the manufacturer's instructions. Each isolate was prepared and EBs purified as described²⁴¹. DNA was extracted from gradient purified EBs using the Promega Wizard Genomic Purification kit according to the manufacturer's protocol²⁴².

For the remaining individuals (118), WGS data were obtained directly from clinical samples. DNA baits spanning the length of the Ct genome were compiled by SureDesign and synthesized by SureSelectXT (Agilent Technologies, UK). Ct DNA extract from clinical samples was quantified and carrier human genomic DNA added to obtain a total of 3µg input for library preparation. DNA was sheared using a Covaris E210 acoustic focusing unit (Christiansen 2014). End-repair, non-templated addition of 3'-A adapter ligation, hybridisation, enrichment PCR and all post- reaction clean-up steps were performed according to the SureSelectXT Illumina Paired-End Sequencing Library protocol (V1.4.1 Sept 2012). All recommended quality control measures were performed between steps.

DNA was sequenced at the Wellcome Trust Sanger Institute using Illumina paired-end technology (Illumina GAII or HiSeq 2000). All 126 sequences passed standard FastQC quality control criteria²⁴³.

3.6.2. Alignment, assembly and filtering by individual genes

Raw fastq files were aligned and assembled using BWA SAMtools²⁴⁴ with A/Har-13 as the reference genome. Variants were called and filtered using BCFtools²⁴⁵ and VCFtools²⁴⁶, with a minimum base quality score of 20 (99% accuracy) and a minimum read depth of 10.

Assembled isolates were combined and used as a database in the command-line version of Basic Local Alignment Search Tool (BLAST+)²⁴⁷. Individual gene sequences from A/Har-13 were used as queries to extract copies successfully sequenced in the isolates. Sequences with more than half missing calls were excluded. MUSCLE algorithm was used for gene alignments²⁴⁸. Alignments were inspected manually using SeaView²⁴⁹ and visualisation were output using Geneious²⁵⁰.

3.6.3. Allele frequency-based signatures of selection

Aligned multi-fasta files for each gene were used as input for Variscan-2.0.3²⁵¹ to calculate Tajima's D, Fu and Li's D* and F* and Fay and Wu's H. RunMode 12 and RunMode 22 were used, sites with less than 50 sequences were not included. Sliding-window analyses were performed over windows of 42 nucleotides with jumps of three nucleotides. All three measures look at the number and frequency of mutations within a population to determine whether they occurred randomly under neutrality or were caused by a form of natural selection. They are based on different methods of estimating the genetic diversity (θ) in a population²⁵². The rationale for using these different metrics is that when combined they are more informative than any one of them alone.

Tajima's D is based on the prevalence of low and intermediate frequency alleles and can detect directional selection (where a single allele is favoured, causing changes in the allele frequency over time) and balancing selection (multiple alleles are maintained at an intermediate frequency in a population). If directional selection is acting, purifying or positive selection, most mutations at a given position will be rare because they are either; deleterious, therefore being driven out of the population (purifying) or one mutation is beneficial, therefore this mutation will increase in frequency at the expense of other mutations (positive). If balancing selection is acting,

two or more mutations at a given position will be present in the population at intermediate frequency. Directional selection is indicated by negative values of D . Balancing selection is indicated by positive values of D .

Tajima's D can be influenced by changes in population structure. If a population has expanded in number, there will be many rare mutations that have not yet been driven out by purifying selection, leading to a negative D -value. Conversely, if a population has decreased in number, there will be fewer rare mutations, leading to a positive D -value. Fu and Li's D^* and F^* are very similar to Tajima's D , however they utilise a sequence from outside the population, to account for the effect of population changes. Therefore, if D and D^*/F^* are positively correlated, this supports the observed mutations being a result of natural selection not population changes.

The above metrics can identify evidence of balancing and directional selection, they cannot distinguish within directional selection between positive and purifying selection. Fay and Wu's H is based on the prevalence of intermediate and high frequency alleles and can detect positive and purifying selection. If purifying selection is acting, few mutations will reach high frequency. If positive selection is acting, beneficial mutations will reach high frequency. Positive selection is indicated by negative values of H . Purifying selection is indicated by positive values of H . Combining these metrics can therefore reliably identify mutations caused by natural selection and determine whether they are under positive, purifying or balancing selection. Calculations for these metrics are detailed below.

Tajima's D compares the average pairwise diversity (π), the average difference between a pair of sequences across all sites, and the number of segregating sites (κ), the number of sites within a population which are polymorphic²⁵³. Tajima's D is calculated from equation 7. For equation 5, x is the frequency of sequences i and j , δ is the number of nucleotide differences per site between them and N is the total number of sequences. For equation 6, n is the number of sequences and i is the number of times a given allele is present.

$$\theta_{\pi} = \sum x_i x_j \delta_{ij} / N \quad (5)$$

$$\theta_{\kappa} = \frac{\kappa}{\sum_{i=1}^{n-1} 1/i} \quad (6)$$

$$D = \theta_{\pi} - \theta_{\kappa} \quad (7)$$

Positive selection increases the frequency of few advantageous mutations, meaning most mutations are kept at a low frequency. In this situation the average difference between pairs of sequences is low but the number of segregating sites is relatively high, thus D will be negative. Purifying selection reduces the frequency of deleterious mutations, meaning mutations are occurring but not becoming common. In this situation again the average difference between pairs of sequences is low and the number of segregating sites is high, thus D will be negative. Balancing selection maintains multiple mutations at intermediate frequencies, this means increased polymorphism at these sites. In this situation the average difference between pairs of sequences is higher but the number of segregating sites remains stable, thus D will be positive. One problem with this measure is that when populations expand mutations are less likely to be lost, meaning the number of segregating sites will increase and D will be negative and the opposite is true for population bottlenecks.

Fu and Li's D^* and F^* are similar to Tajima's D but they distinguish between 'old' and 'young' mutations, and therefore are less sensitive to population changes²⁵⁴. 'Old' mutations are found on internal branches of the genealogy and 'young' mutations are found on external, more recent branches. D^* is calculated from equation 8, essentially total number of mutations minus number of singletons, mutations only occurring once in the population. F^* is calculated from equation 9, average pairwise diversity minus number of singletons. For equations 8 and 9, n is the number of sequences, η is the total number of mutations, α_n is the denominator from equation 6 and η_s is the number of singletons.

$$D^* = (n/n - 1) \eta - \alpha_n \eta_s \quad (8)$$

$$F^* = \theta_\pi - \frac{n-1}{n} \eta_s \quad (9)$$

Similarly to Tajima's D positive and purifying selection both bring about an excess of rare mutations, making D^* and F^* negative. In contrast balancing selection will maintain older mutations, reducing the number of 'young' mutations making D^* and F^* positive. By comparing Tajima's D with Fu and Li's F^* we can determine the influence of singleton mutations, an F^* value more negative than its corresponding D value would indicate that most mutations are singletons. As described above, negative D , D^* and F^*

values can be indicative of both positive and purifying selection. Fay and Wu's H can be used to determine which form of natural selection is occurring, or more accurately which is the dominant selection pressure.

Fay and Wu's H utilises an outgroup sequence (A/Har-13) and focusses on the difference between intermediate and high frequency alleles rather than intermediate and low frequency alleles as the previous metrics do²⁵⁵. It is calculated from equation 12. Where n is the number of sequences, i is the number of times an allele occurs and ξ_i is the number of differences from the ancestral outgroup per allele.

$$\theta_{\pi} = 2/n(n-1) \sum_{i=1}^{n-1} i(n-i)\xi_i \quad (10)$$

$$\theta_H = 2/n(n-1) \sum_{i=1}^{n-1} i^2 \xi_i \quad (11)$$

$$H = \theta_{\pi} - \theta_H \quad (12)$$

H is heavily influenced by high frequency alleles because of the i^2 component in equation 11. Purifying selection keeps alleles from becoming common, therefore there will be few high frequency alleles and H will be positive. Positive selection causes alleles to rise to fixation thus increase in frequency, therefore H will be negative.

3.6.4. Haplotype-based signatures of selection

These metrics utilising allele frequencies were complemented by a haplotype based method, integrated haplotype score (iHS). iHS was designed to identify alleles at intermediate frequencies that are being driven towards fixation (positive selection) or possibly balancing selection²⁵⁶. iHS looks at SNPs individually and their proximity along the chromosome to other SNPs, comparing a user-defined ancestral sequence to the derived sequence. In this study, the ancestral sequence chosen is A/Har-13, a trachoma-isolate collected in the 1950s. iHS is able to identify regions along a chromosome with relatively recent evidence of positive selection, by comparing the acquisition or loss of SNPs, and those in close proximity, between the ancestral and

derived sequences. A significantly positive or negative iHS, determined by comparing to all SNPs on the chromosome, is indicative of positive selection.

Extended haplotype homozygosity (EHH) is a measure of distance calculated around SNPs on a given haplotype, if the haplotype is all the same EHH will be one and if it is all different EHH will be zero. Integrated haplotype homozygosity (iHH) is calculated as the decrease or decay in EHH as you move further from a given SNP, both on the ancestral sequence and the derived sequence. In our studies the ancestral allele was defined as the allele in A/Har-13 and the derived allele was defined by variants in the Ct isolates from Guinea-Bissau. The unstandardised iHS is calculated from equation 13. Where iHHA and iHHD are the iHH from the ancestral and derived sequences respectively.

$$\text{unstandardised iHS} = \ln\left(\frac{iHHA}{iHHD}\right) \quad (13)$$

Unstandardised iHS is then standardised based on the genome-wide mean and variance of iHS for SNPs with the same allele frequency. If multiple SNPs with evidence of selection are within a defined distance of each other, EHH greater than 0.05, they can be combined to indicate windows under selection. Values significantly less than one indicate longer haplotypes on the derived sequence, indicative of a selective sweep driving an allele towards fixation before further mutations have arisen. Values significantly greater than one indicate longer haplotypes in the ancestral sequence, this is also a sign of selection. Selection could now be favouring the ancestral allele or ancestral alleles around the favoured site could be hitchhiking with it. In our analysis we will use the absolute iHS, using significantly positive values as indicators of selection.

Short read data from the 126 ocular Ct isolates were mapped against Ct A/HAR-13 using SAMtools²⁴⁵. Non-polymorphic sites were removed. SNPs with a minor allele frequency (MAF) less than 0.05 and more than 25 % missing calls were excluded. Alleles were defined as ancestral (0) or derived (1) based on the ancestral isolate A/Har-13. iHS incorporate distance between SNPs to identify signatures of selection and cannot be calculated over sites with missing calls, for this reason imputation was used to classify missing calls. A simple genetic distance-based imputation was used. Genome-wide pairwise nucleotide diversity was calculated for each pair of sequences. For each missing call within a sequence, all sequences with calls at those sites were assigned a

score based on the pairwise diversity between them and the sequence with a missing call. These scores were summed at each site for ancestral and derived calls respectively and divided by the number of respective calls. The state with the lower score was used to define the missing call (Figure 3.2). iHS was calculated using the R package rehh, WHAMM (<http://coruscant.itmat.upenn.edu/whamm/ihs.html>) and selscan²⁵⁷. Scores were standardised as described above after binning MAF into 40 discrete bins of size 0.025.

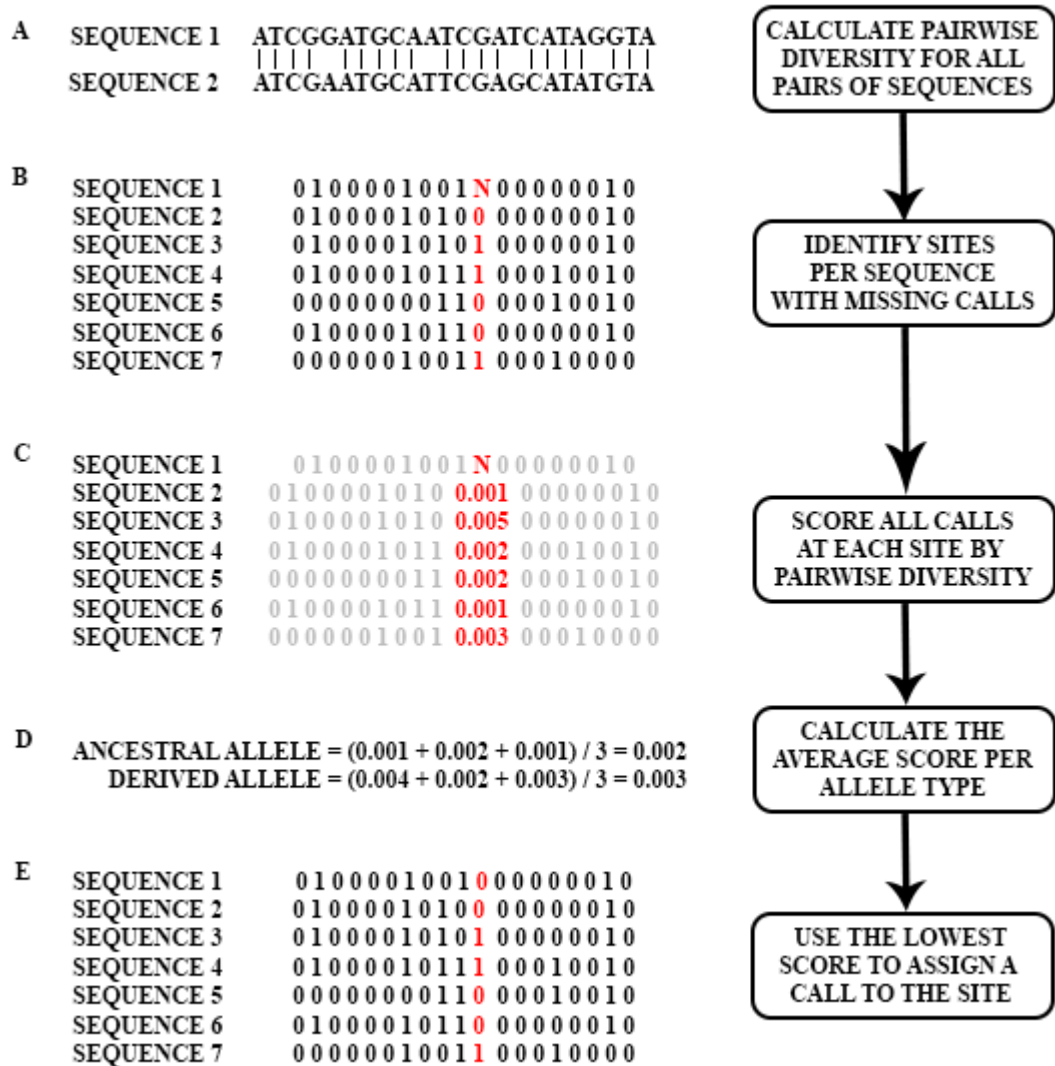


Figure 3.2. Genetic distance-based imputation of missing calls.

A) Genome-wide pairwise diversity was calculated for each pairs of sequences. B) Each sequence was examined for missing calls. C) For each site with a missing call for each sequence, calls at each site for assigned a score based on their sequences pairwise diversity when compared to the sequence with the missing call. D) These scores were

summed for the ancestral allele (0) and the derived allele (1) respectively and divided by the number of each allele to get an average score per allele type. E) If the average ancestral allele score was lower than the equivalent derived allele score the missing call was classed as ancestral (0) and vice versa, because a lower score means the sequences are more similar.

3.7. Cell biology using cell-culture models

3.7.1. Mammalian cell-culture and Ct infection

Cell lines were cultured in 75 cm² aerated flasks at 37 °C with 5 % CO₂ and split every 2 days at approximately 80 % confluence. HeLa cells were cultured in culture medium A and seeded at a density of 2x10⁶ cells per ml. For microscopy, cells were seeded at a density of 7x10⁴ cells per ml at approximately 50 % confluence onto glass coverslips in 24-well plates.

Ct-LGV2 infections were performed 24 hours after seeding of cells onto coverslips, -80 °C stocks were diluted to one infection forming units (IFU) in infection medium and centrifuged at room temperature for 10 minutes at 1000 RPM. After incubation at 37 °C five % CO₂ for 80 minutes infection medium was exchanged for the relevant culture medium and cells were incubated at 37 °C, 5 % CO₂ until fixation.

3.7.2. GFP-construct transfection

Green fluorescent protein (GFP)-construct transfections were performed 24 hours after seeding of cells onto coverslips. Constructs were diluted to 150 ng per 50 µl transfection medium and incubated at room temperature for fifteen minutes, followed by 1/10 dilution into culture medium onto cells and incubation at 37 °C, 5 % CO₂ until fixation.

3.7.3. Fixation and antibody staining for IF microscopy

Cells were fixed with either 4% paraformaldehyde (PFA) at room temperature for 30 minutes, followed by addition of 50 mM ammonium chloride, or methanol at -20 °C for 5 minutes. Fixed coverslips were washed 3 times with PBS and cells were

permeabilised with either 1:1 methanol/ethanol or PBS with 0.05 % Triton X-100 at 4 °C for 15 minutes. After 2 washes coverslips were blocked with 1 or 10 % BSA. After 2 further washes primary antibody in PBS with 1 % BSA was added and coverslips were incubated at room temperature (Table 3.13). After 2 washes coverslips were stained with Hoechst 33258 (1/5000) and relevant secondary antibody (1/200) in PBS with one % BSA and incubated at room temperature for 30 minutes. After 2 final washes coverslips were mounted onto slides using MolWiol and stored at 4 °C for at least 24 hours before examination.

Table 3.13. Primary antibodies for IF microscopy.

PRIMARY ANTIBODY TARGET	DILUTION	HOST SPECIES
CT442	1:200	Rabbit
CALRETICULIN	1:1000	Rabbit

3.7.4. Co-Immunoprecipitation of CT442-GFP from HeLa cells

HeLa cells were seeded and cultured as above but in 150 mm by 20 mm plates for a total of 8×10^6 cells per condition, conditions tested were GFP or CT442-GFP transfected with and without Ct-LGV2 infection. Twenty-four hours post-transfection and infection cells were washed twice with Hank's buffered saline solution (HBSS) and incubated on ice for 5 minutes with co-immunoprecipitation (Co-IP) buffer. Cells were scraped from the flasks and vortexed for 1 minute, followed by incubation on a roller at 4 °C for 30 minutes while being vortexed every 5 minutes. Cells were lysed by ten passages through a 23G needle and supernatant was collected by centrifugation at four °C for ten minutes at 13000 RPM. Five µl Protein-G beads (Thermo Fisher Scientific) were added per ml of lysate, followed by incubation on a roller at 4 °C for 2 hours. The beads and non-specifically bound proteins were removed using a magnetic bar and 1 µl anti-GFP antibody was added per ml of lysate, followed by incubation on a roller at 4 °C overnight. Five µl new Protein-G beads were added per ml of lysate, followed by incubation on a roller at 4 °C for 3 hours. The beads and specifically bound proteins were collected using a magnetic bar and the beads were washed 3 times with PBS with 0.1% NP40. After 3 washes in PBS the beads were split in two and stored at -20 °C.

Elution buffer B was added to half of the beads and vortexed for 3 minutes, the eluate was diluted 1:1 with Laemmli buffer and incubated at 95°C for 10 minutes. Samples were blotted as described previously and bands were visualised using anti-GFP antibody to confirm precipitation of GFP and CT442-GFP from cultures.

3.7.5. Liquid chromatography-mass spectrometry

The remaining half of the beads were analysed by liquid chromatography-mass spectrometry as described previously⁸⁷. Briefly, bead-coupled proteins were trypsin-digested overnight at 37°C in twenty µl 25 mM NH₄HCO₃. Digests were analysed using a LTQ Velos Orbitrap (Thermo Fisher Scientific) coupled to a nano-LC Ultimate 3000 RSLCnano system (Thermo Fisher Scientific). Peptides were analysed in the orbitrap in full ion scan mode at a resolution of 30,000 (at m/z 400) and with a mass range of m/z 400–1800. Mass spectrometry data were searched against the NCBI *Homo sapiens* and the SwissProt *Chlamydia trachomatis* databases. False discovery rate (FDR) was calculated using the reversed database approach with a 1 % filter.

3.7.6. Filtering of mass spectrometry hits

Hits from mass spectrometry were grouped into 7 categories; GFP only, ubiquitous, uninfected only, infected only, CT442-uninfected, CT442-ubiquitous and CT442-infected (Table 3.14). Hits from groups 1 to 5 were excluded, for Ct hits group 6 was also excluded. Hits were ranked based on the score from group 7, those below 50 were excluded. Overlapping hits common to five or more inclusion membrane proteins (Incs) from a large-scale affinity-based mass-spectrometry analysis of Inc-host interactions were also excluded⁹⁶. Overlap with CT442 proteins identified in the above study, Ct-inclusion enriched proteins²⁵⁸ and lipid-droplet enriched proteins from Ct-infected cells⁹³ were included but not used to filter.

Table 3.14. Grouping of proteins identified by mass spectrometry.

Proteins were grouped for filtering of targets. A '+' indicates proteins that were present in the list of targets from that experiment.

GROUP	UNINFECTED		INFECTED		DESCRIPTION
	GFP	CT442-GFP	GFP	CT442-GFP	
1	+	-	+	-	GFP only
2	+	+	+	+	Ubiquitous
3	+	+	-	-	Uninfected only
4	-	-	+	+	Infected only
5	-	+	-	-	CT442-uninfected
6	-	+	-	+	CT442-ubiquitous
7	-	-	-	+	CT442-infected

3.7.7. Pathway-enrichment analysis

Pathway enrichment analysis was performed using all filtered proteins, either including or excluding proteins which were common to 5 or more Incs as described above. The comparator dataset for enrichment analysis was *Homo sapiens* proteome, NCBI taxon-id 9606. Bonferroni correction was used to correct for multiple comparisons, corrected p-values ≤ 0.05 were considered significant. The Database for Annotation, Visualization and Integrated Discovery (DAVID) bioinformatics resource²⁵⁹ was used to determine enrichment of GO-terms²⁶⁰ and Pfam²⁶¹ and InterPro protein domains²⁶². Enriched pathways were determined through KEGG²⁶³ and Reactome²⁶⁴. Protein-protein interactions were found and scored using STRING²⁶⁵, interactions with an overall score less than 0.5 were excluded. Visualisations of protein-protein interactions were produced using R package igraph. Spheres contain filtered hits. Size of protein-spheres represents mass spectrometry scores for each protein, increased size reflects higher score. Connecting lines represent interactions. Thickness of connecting lines represent

STRING interaction scores, increased thickness reflects higher scores. Ct-proteins were manually annotated with published information.

4. Association between antibody response and longitudinal ocular infection

4.1. Introduction

4.1.1. Natural history of trachoma

Ct has consistently been isolated from the ocular conjunctiva of individuals with clinical signs of trachoma⁴², and had been shown through direct inoculation to cause trachomatous disease^{43, 44}. In trachoma-endemic communities ocular Ct infection has been consistently shown to associate with the presence and severity of active trachomatous disease (TF and/or TI)^{266, 267}. Ct-infected individuals without active disease still often have sub-clinical minor signs of disease and live in households where active disease is present²⁶⁷. In households and communities with increased prevalence and load of ocular Ct, active disease is also more common²⁶⁷⁻²⁶⁹. Individuals with severe disease (TI) have higher Ct conjunctival loads and prolonged episodes of infection²⁶⁸.

Active disease is primarily found in children under the age of ten years²⁷⁰. Disease prevalence declines from its peak in pre-school children (one to four years old) to older children and teens (five to fourteen years old and from this group to adults (fifteen years or older)^{8, 266, 271}. This can partly be explained by reduced exposure to Ct with increasing age, which is presumed to be due to development of immunity^{270, 272}. This reduced prevalence observed in cross-sectional surveys is also partly due to duration of infections and active disease episodes which both decline with age. Median infectious episode duration drops from approximately four weeks in young children to two weeks and less than two weeks in older children and young adults respectively⁸. Similarly, median active disease episode duration falls from thirteen weeks to five and two weeks in the older age groups. Modelling data suggests the median durations of both infection and active disease may be longer, particularly in pre-school children, however the decrease in length of episodes is consistent⁹. This suggests a partial immunity to Ct develops with increasing age in endemic communities.

4.1.2. A balance of protective and pathogenic immune responses

Conversely to active disease, prevalence of trachomatous scarring, trichiasis and corneal opacities increase with age²⁷¹. Persistent or chronic infection is associated with progression of these pathologies²⁷³. This presents a paradoxical situation where immune

responses are necessary in resolving Ct infection but can promote or at least do not prevent later pathologies²⁷⁴. Conjunctival Ct infection induces a strong pro-inflammatory response marked by production of cytokines such as interleukin-1 alpha (IL-1 α), IL-6, IL-8 and granulocyte macrophage colony-stimulating factor (GMCSF)^{117, 118}. This innate immune response also promotes recruitment of neutrophils, macrophages and NK-cells¹¹⁹⁻¹²¹. NK cells are also important during Ct infection through early secretion of IFN γ and activation of Th1 CD4⁺ T-cells²⁷⁵. Cell-mediated immune responses have been extensively studied in chlamydial infections, proliferation of CD4⁺ T-cells and production of interferon-gamma (IFN γ) has been implicated in successful resolution of infection in animal models and human infections^{125-127, 142, 186}. Lymphoproliferative and cytolytic responses to Ct antigens including whole EBs and MOMP are important in the resolution of infection¹²⁴⁻¹²⁶. Recent work using a non-human primate model of trachoma has suggested CD8⁺ T-cells may also have a role in Ct clearance²⁰⁰.

The role of antibodies in Ct infection and subsequent disease has been the subject of decades of research as Ct infection stimulates and influx of B-cells into the conjunctiva¹¹⁹. These are the primary constituent of conjunctival follicles and stimulate hypergammaglobulinaemia¹³⁰. Neutralising antibodies against Ct have been demonstrated in animal models^{131, 132} and *in vitro*¹³³⁻¹³⁵ and B-cell promotion of T-cell mediated Ct clearance is important in some murine models²⁷⁶⁻²⁷⁸. A recent study in non-human primates showed increased antibody responses against a panel of Ct antigens were heightened in animals that were partially protected upon ocular re-challenge with Ct¹³⁶. In contrast, human studies find anti-Ct antibodies associated with both active disease and progressive scarring^{129, 266, 279}, with the highest titres found in cases of severe disease¹³⁰. A longitudinal study in The Gambia found that higher IgG responses against the immunodominant MOMP were associated with higher rates of infection and higher titres increasing the associated risk²⁸⁰. Ocular Ct infection clearly induces a strong humoral immune response but its role in protection or pathology *in vivo* is unclear.

4.1.3. Proteome-wide antibody profiling

Screening whole-proteome arrays as discussed in chapter 1.4 has become a popular method to elucidate antibody responses and improve understanding of their role in infection and disease²⁸¹. Extensive use of these methods has highlighted some common themes in humoral immune targets of human pathogens including; functions in protein binding and catalytic activities, early/late or late expression stage and membrane localisation²⁸². A recent study on *Plasmodium falciparum* used a panel of 856 previously identified parasite antigens to screen samples from a Ugandan population with well characterised *P. falciparum* infection in the last year and malaria exposure in the previous eight years²⁸³. This longitudinal information allowed them to identify current antibody responses that were predictive of most recent exposure and frequency of exposure in the preceding twelve months. Longitudinal studies have previously been utilised to examine the impact of tear and serum antibody responses, against Ct-EB and the Ct antigen MOMP, on Ct infection²⁸⁰. A similar methodology was applied to elucidate the role of Ct genome-wide antibody responses in longitudinal ocular Ct infection in a trachoma-endemic setting.

4.1.4. Study design and arrayed sera

The study was performed as described in chapter 3.4.3 and in Faal *et al*¹⁴². Briefly, 345 children aged 4 to 15 were recruited from nine villages in The Gambia where rapid assessment of school-age children found greater than 20 % active trachoma. At baseline and each fortnightly visit for 6 months, children were examined for clinical signs of active trachoma. Conjunctival swabs were collected in duplicate to test for Ct infection and tear fluid was collected for cytokine-based and serological assays. Blood samples were collected at baseline and cessation of the study, a total of 186 samples were collected. Conjunctival expression and production of cytokines in this cohort have been previously examined¹²⁸. Both expression and production of pro-inflammatory cytokines, predominantly IFN γ and TNF α , was enhanced in individuals with current ocular Ct infection. This pro-inflammatory response was also associated with acquisition and slower resolution of Ct infection. This study attempted to characterise the humoral immune response in these individuals.

The number of samples to be screened on the array was limited due to resource limitations. Samples were selected to cover a range of phenotypes, including; recent infection, multiple episodes of infection, lack of infection and/or clinical disease. In total 90 sera from baseline and 33 sera from the end of the study were selected. These were screened against a previously published array incorporating 894 genomic ORFs from serovar D Ct (chapter 3.2.1)²⁰⁵. Individual antigen fluorescence intensities were assumed to be proportional to the presence of antigen-specific antibodies in the sera. The complete profile of responses for each serum sample was used to look for global changes in antibody recognition and its association with frequency and duration of Ct infection. Differential recognition of individual antigens was also examined. In the analysis presented only the 90 prospective baseline sera were included in order to identify which antigens were the focus of current or established immune responses associated with protection from acquisition of further Ct infection and disease.

4.2. Results

4.2.1. Protection from Ocular Ct infection

A number of outcome measures were considered to reflect immunity. The simplest comparison being individuals free of infection during the six months against those with any infection, however the latter covered individuals with one to five infectious episodes who were clearly resolving infection differently. Separating this group into one or more episodes of infection was also inappropriate as single episodes lasted between two and 26 weeks. Therefore frequency and duration of infection episodes were combined to define acquired immunity and to split individuals into groups for analysis. Initially three groups of individuals were compared. Those with no infections, those with short duration infections (two weeks or less) and those with long duration infections (longer than two weeks). In order to increase the power of analysis, those with no or short duration infections were combined (protected-immune) and compared to those with long duration infections (susceptible – non-immune). Short and long duration infections were defined by the median duration of infectious episodes for the 90 individuals with sera. The median duration of infection was two weeks or one visit (Table 4.1). This comparison compares individuals who were either protected from

infection or resolved quickly against those with either long duration or frequent infection episodes (Figure 4.1), which reflects the variation in immunity seen in this study and other trachoma-endemic communities.

Protected and susceptible individuals were comparable by age and gender. Village membership was significantly different but both protected and susceptible individuals were present in four of six villages. The demographic similarity of immune and non-immune individuals meant that history of ocular Ct exposure was assumed to be similar.

Table 4.1. Patient demographics in protected and susceptible groups.

Associations with protection or susceptibility to infection were determined using a generalised linear model. For age, bracketed numbers are the 95 % confidence intervals around the median. For gender and village, bracketed numbers are the percentage.

	PROTECTED (IMMUNE)	SUSCEPTIBLE (NON-IMMUNE)	P-VALUE
NUMBER	60	30	NA
AGE IN YEARS (95% CI)	9.00 (2.00-11.53)	8.00 (1.73-12.00)	0.413
FEMALE (N [%])	24 (40.00)	13 (43.33)	0.762
VILLAGE (N [%])			< 0.001
- 1	7 (11.67)	19 (63.33)	
- 2	2 (3.33)	0 (0.00)	
- 4	5 (8.33)	1 (3.33)	
- 6	11 (18.33)	3 (10.00)	
- 7	27 (45.00)	7 (23.33)	
- 8	8 (13.33)	0 (0.00)	

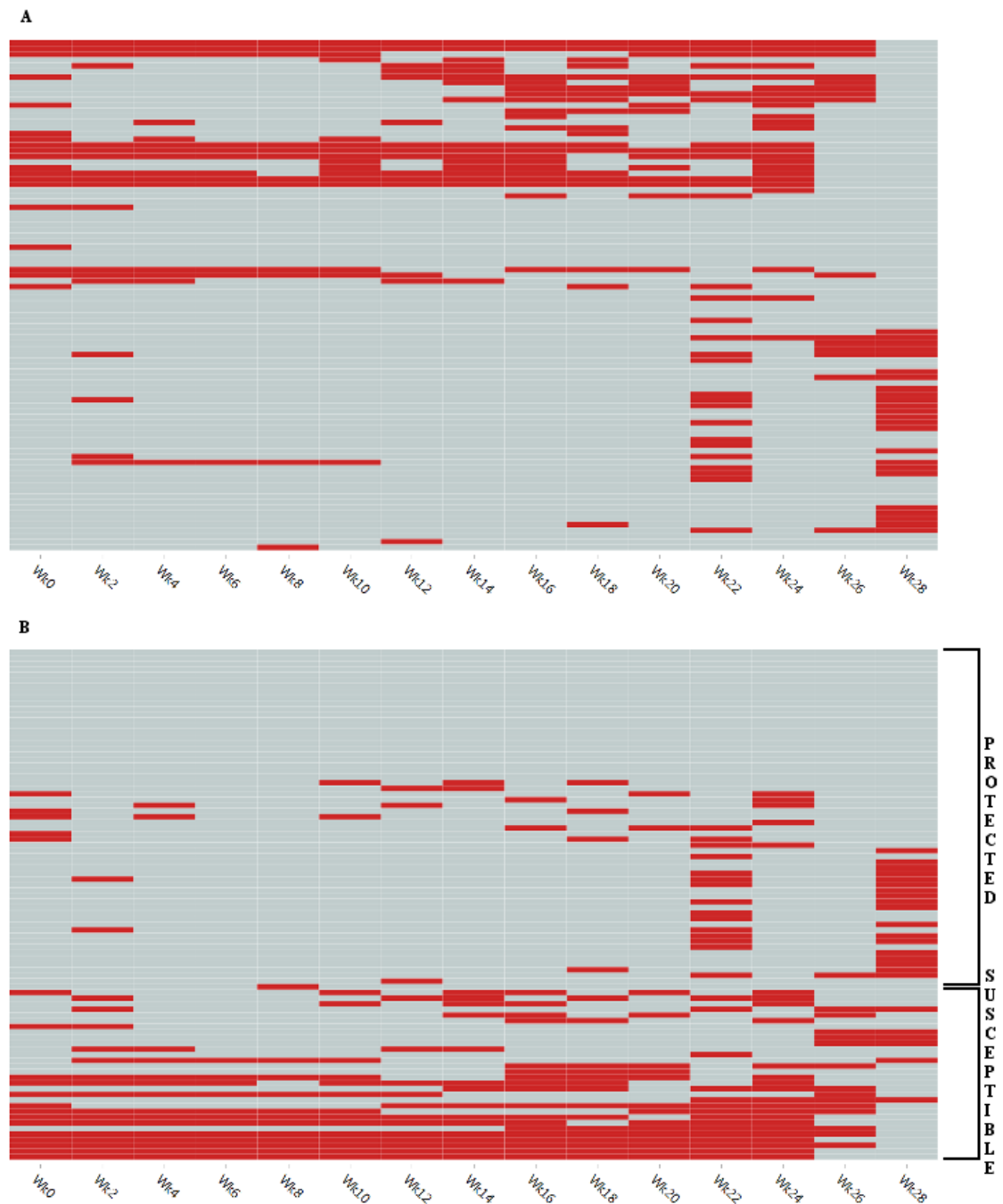


Figure 4.1. Visualisation of longitudinal evidence of ocular Ct infection in 90 individuals during the six-month study.

Visits every two-weeks are indicated on the horizontal axis, the 90 individuals are separated on the vertical axis. Red rectangles are visits where an individual's conjunctival swab was Ct-positive by 16S rRNA QPCR, grey rectangles are Ct-negative visits. A) Individuals are sorted by ID. B) Individuals are sorted by median duration of infection episodes, protected and susceptible groupings are indicated on the right-hand side.

4.2.2. Array normalisation and filtering

The raw fluorescence intensity data from the microarray was strongly positively skewed, $p\text{-value} < 0.001$ (Figure 4.2A). In total 90 serum samples from the beginning of the study and 33 serum samples from the end tested on 894 antigens were included. Negative values were classified as missing calls (NA) as they were strongly separate from the distribution of the remaining data.

Transformation and normalisation of the array was assessed using relevant rank deviation (RRD)²⁰⁶. The data was inverse hyperbolic sine transformed, which reduced RRD across all samples (Figure 4.2B). Mean-centring and trimmed means of 2 and 10 % were tested to normalise the data. The two trimmed mean methods were indistinguishable, mean-centring reduced the RRD compared to both (Figure 4.2C).

Antibody responses were filtered by excluding any with mean fluorescence intensity lower than the global mean across all antigens, this removed infrequently recognised antigens. The 30 serum samples from the end of the cohort were excluded because of the small group size and an intention to focus on prospective responses. The normalised data from 90 serum samples tested on 441 antigens remained positively skewed ($p\text{-value} 0.001$) as a result non-parametric tests of significance were used throughout (Figure 4.2D).

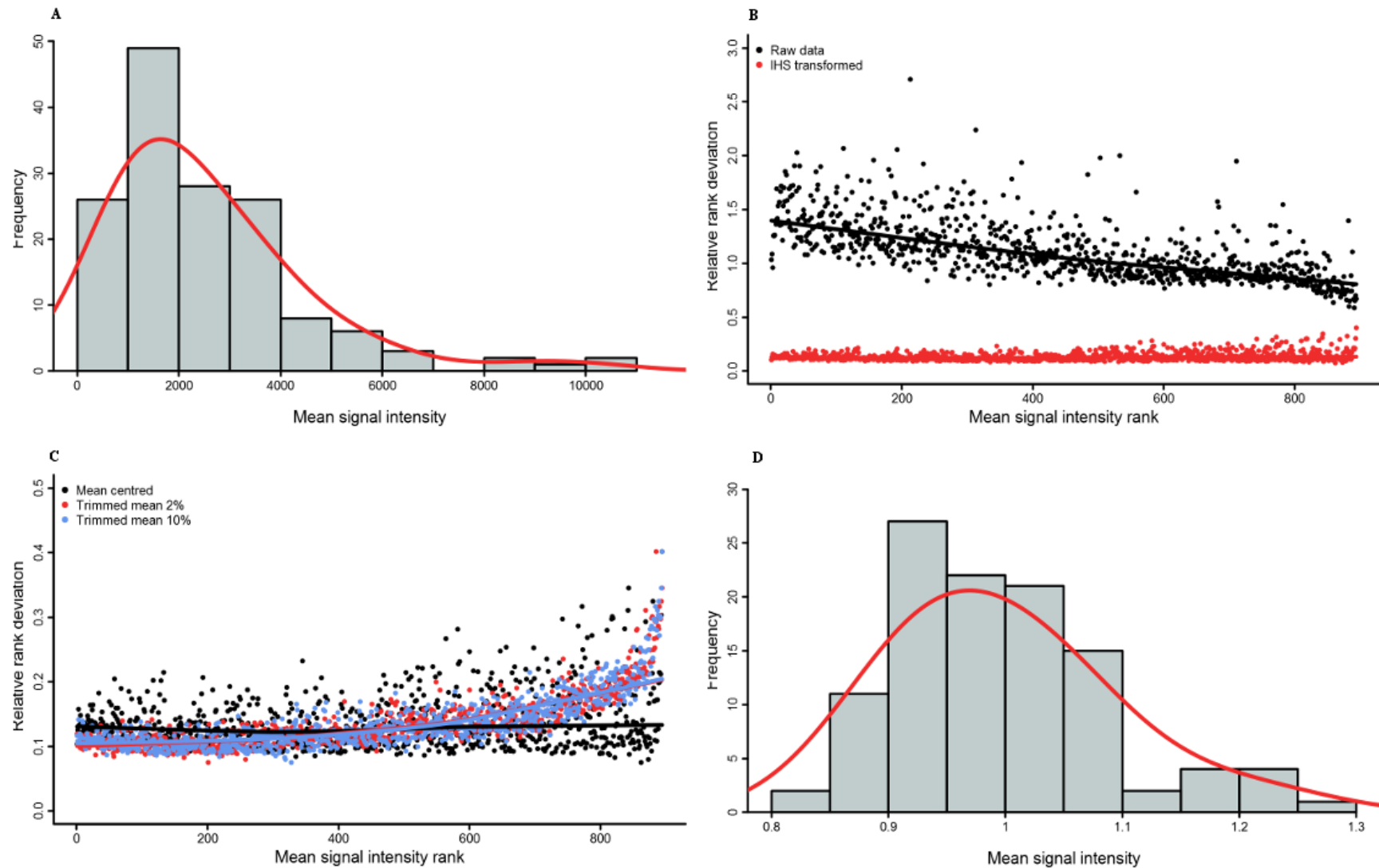


Figure 4.2. Transformation and normalisation of the raw array data.

A) Histogram showing positively skewed distribution of the raw data. B) Significantly reduced deviation (RRD) in the data after inverse hyperbolic-sine transformation (red) compared with the raw data (black). C) Reduced deviation in the data after mean-centring (black) compared with normalisation by trimmed-mean at 2 % (red) and 10 % (blue). D). Histogram of normalised data still showing a slight positive skew.

The normalised continuous data was used where its distribution permitted the appropriate analysis, for methods which required categorical data samples were divided into groups on a per antigen basis. Division of the data into groups was assessed using average silhouette width, which tests within and between group variance to determine the degree of separation of individuals. Grouping methods tested were; k-means clustering, k-medoids clustering, fuzzy c-means clustering, hierarchical clustering or mixture modelling.

Binary classification into positive and negative samples per antigen performed best and was empirically better than all other cluster sizes up to ten, p-value < 0.001 (Appendix Figure 1). Hierarchical clustering was the worst performer for a cluster size of two across all 441 antigens. The remaining methods while not performing well (median silhouette width average = 0.6) were equally good at splitting the data (Figure 4.3). Instead of using one method for classifying responses as positive, silhouette score was determined for individual antigens using each clustering method. For each antigen the clustering method with the highest silhouette score was used to call responses positive or negative. This method, referred to as 'Best', was significantly better than using one method for all antigens, p-value < 0.001 (Figure 4.3).

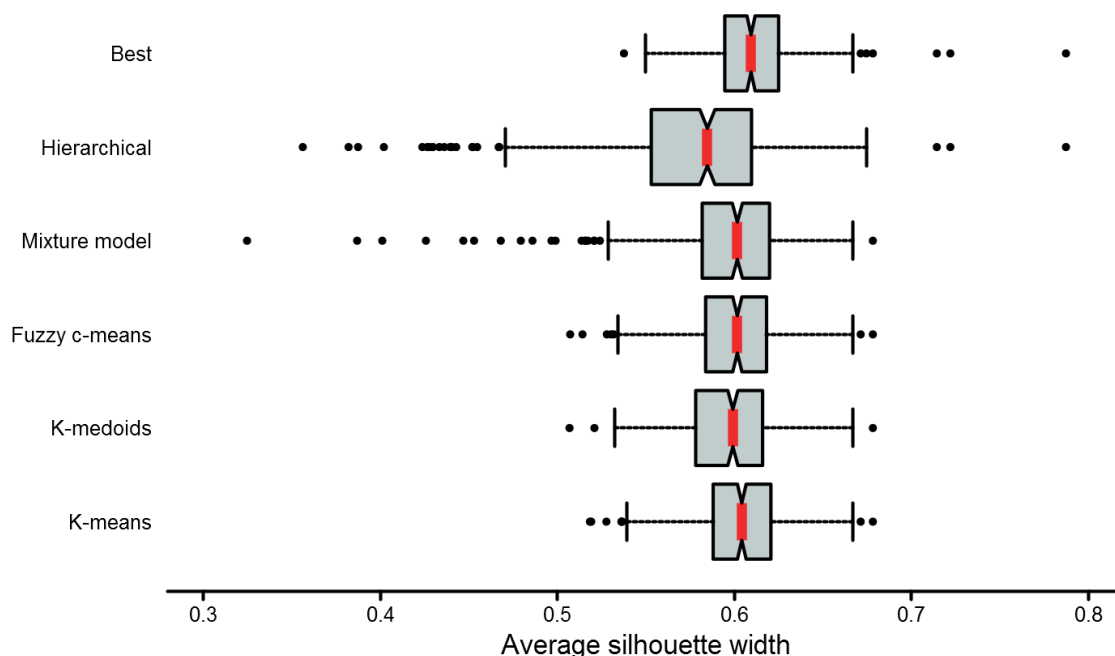


Figure 4.3. Average silhouette widths for clustering method trialled for all 441 antigens.

'Best' method had the highest median across all antigens. Clustering methods are detailed on the left-hand side. Red lines indicate the median. Notches were calculated as median $\pm 1.57 \times \text{IQR} / \sqrt{n}$, where IQR is the interquartile range and n is the number of samples. The whiskers were calculated by adding 1.5 times the IQR to the 75 percentile and subtracting 1.5 times the IQR from the 25 percentile. Dots are outliers.

4.2.3. More focussed global antibody responses are associated with protection

Comparing 441 filtered antigens with the complete panel of 894 present on the microarray there was no difference in gene expression stage or peak, p-values of 0.772 and 0.498, and no difference in predicted localisation, p-value 0.390.

The global responses were examined in protected and susceptible groups. Breadth, defined as the number of antigens an individual had a positive response to, was highly variable. There was a non-significant trend towards lower breadth in protected individuals, p-value 0.088 (Figure 4.4).

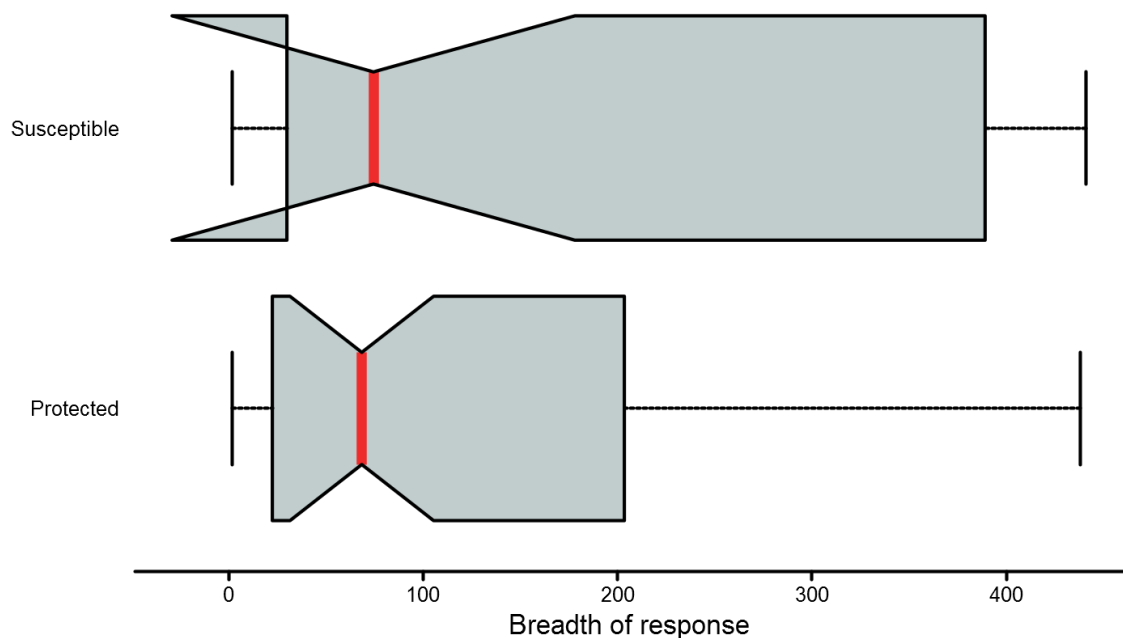


Figure 4.4. Increased breadth of responses in susceptible individuals.

Boxplots of breadth of responses (x-axis) in protected and susceptible individuals (y-axis). Breadth was measured as the number of antigens an individual had a positive response to.

Global antibody profiles were examined further by adaptation of measures of diversity typically applied in ecology and now in meta-genomics²⁸⁴, these incorporate breadth and relative abundance of species²¹⁰. In this context each individual was considered as a separate population of antibody responses (species), the relative strength of which

equated to abundance (chapter 3.2.6). The diversity per individual was measured using three different methods, Hill numbers, Shannon's diversity indices and Simpson's diversity indices because they often produce different results²¹¹. A high number indicates high diversity or an evenness of antibody responses, meaning responses are targeted equally across multiple antigens. Hill numbers also incorporate order of diversity. Increasing this number puts greater importance on the most abundant antibody responses. Hill numbers stable across changing order of diversity indicate an even population of responses, declining numbers as order of diversity increases indicate an uneven population dominated by a few commonly recognised antigens.

Hill numbers dropped significantly with increasing order of diversity in both groups, which means individual's responses were highly uneven and therefore focussed on a few select antigens. Hill numbers were similar across all orders of diversity between the protected and susceptible groups (Table 4.2). Both Shannon's and Simpson's diversity indices were higher in susceptible individuals, p-values of 0.080 and 0.024. This demonstrates less focussed antibody responses in susceptible individuals, as suggested by breadth (Figure 4.5).

Table 4.2. Uneven antibody responses in protected and susceptible individuals, as determined by Hill numbers.

Associations were determined using a generalised linear model.

ORDER OF DIVERSITY	PROTECTED MEDIAN (IQR)	SUSCEPTIBLE MEDIAN (IQR)	P-VALUE
2	415.21 (414.65-415.69)	415.12 (413.99-415.76)	0.472
3	20.38 (20.36-20.39)	20.37 (20.35-20.39)	0.472
4	7.46 (7.46-7.46)	7.46 (7.45-7.46)	0.472
5	4.51 (4.51-4.52)	4.51 (4.51-4.52)	0.472

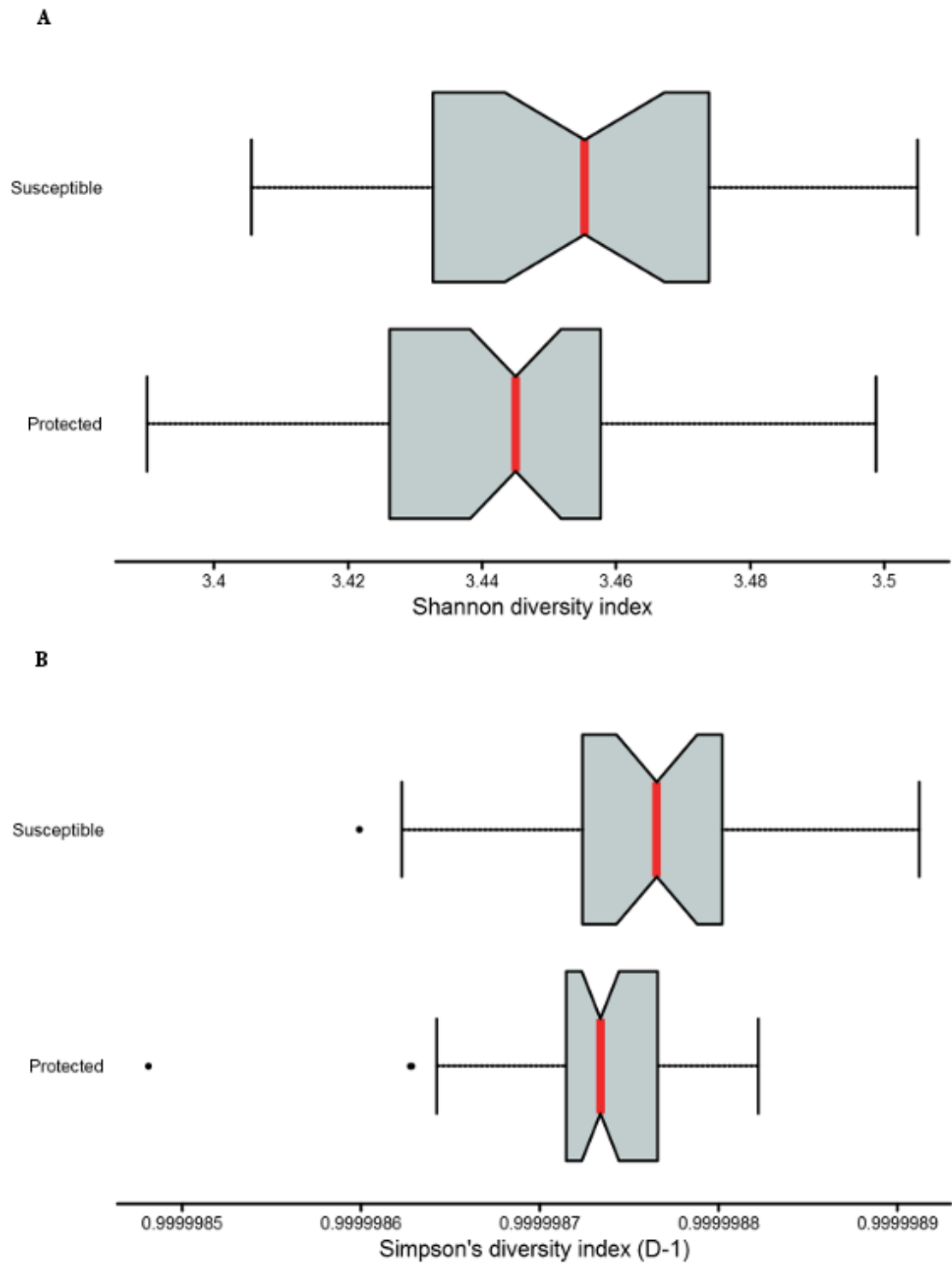


Figure 4.5. Increased diversity and less focussed responses in susceptible individuals.

Diversity was measured using A) Shannon's diversity index and B) Simpson's diversity index.

4.2.4. Individual antibody responses are associated with a lack of protection

Association of responses with individual antigens and infection outcome was determined using a generalised linear model adjusting for age, gender and village of residence. Forty-two antigens were identified as targets of differential antibody responses between protected and susceptible individuals, p-values ≤ 0.05 (Table 4.3). Higher responses to all of these antigens associated with lack of protection from infection. When baseline Ct infection status was included in the model 19/42 antigens remained significantly associated, a further 19 were associated at a significance level of 0.1. The association of higher responses to these antigens with susceptibility to infection was not due to current Ct infection. Despite variation in the significance of the association they were equally good at predicting whether an individual was protected or susceptible, with AUC's ranging from 0.73 to 0.80.

Table 4.3. Forty-two differentially recognised antigens between protected and susceptible individuals.

Univariate associations were determined using a generalised linear model. Variables were resampled 10,000 times and remodelled to determine permuted p-values (P*). Odds ratios and confidence intervals were calculated for an increase of half the range per antigen, rather than one unit (chapter 3.2.5). CT⁺ is the p-value after adjusting for baseline infection status.

ANTIGEN	P-VALUE	P*	T	SE(T)	OR*	95% CI*	AUC	CT ⁺
CT545	0.001	< 0.001	8.38	2.63	9.05	2.57-39.71	0.78	0.003
CT118	0.003	0.002	6.40	2.18	8.18	2.19-37.48	0.78	0.008
CT123	0.004	0.001	6.43	2.21	8.7	2.23-43.16	0.76	0.006
CT541	0.006	0.004	6.09	2.21	6.74	1.85-28.71	0.78	0.012
CT664	0.009	0.008	5.76	2.20	6.36	1.71-28.27	0.80	0.015
CT119	0.009	0.007	5.19	1.98	4.35	1.51-14.00	0.76	0.019
CT381	0.010	0.009	5.03	1.96	4.56	1.48-15.49	0.76	0.021
CT584	0.012	0.010	6.90	2.75	6.3	1.62-29.64	0.76	0.022
CT284	0.012	0.009	6.34	2.53	4.83	1.51-18.22	0.74	0.035
CT502	0.014	0.012	6.46	2.62	5	1.50-19.93	0.80	0.018

CT051	0.017	0.015	5.06	2.11	5.44	1.47-24.26	0.74	0.027
CT592	0.019	0.018	4.14	1.77	3.6	1.28-11.23	0.75	0.038
CT073	0.020	0.017	6.01	2.58	6.91	1.47-39.52	0.77	0.015
CT078	0.020	0.018	6.07	2.61	4.88	1.38-20.58	0.74	0.044
CT023	0.022	0.019	6.60	2.88	7.61	1.56-51.76	0.77	0.038
CT223	0.022	0.020	5.93	2.59	4.11	1.31-15.14	0.76	0.043
CT795	0.024	0.018	3.91	1.73	3.67	1.23-11.93	0.76	0.036
CT875	0.028	0.030	5.14	2.34	4.21	1.22-16.47	0.73	0.070
CT181	0.030	0.024	4.32	1.99	4.05	1.20-15.42	0.75	0.067
CT694	0.031	0.031	3.91	1.81	3.68	1.17-12.78	0.74	0.048
CT021	0.034	0.033	5.45	2.57	4.72	1.19-21.77	0.77	0.074
CT813	0.035	0.033	3.50	1.66	3.13	1.12-9.57	0.75	0.064
CT841	0.036	0.035	3.77	1.80	3.15	1.12-9.78	0.74	0.074
CT017	0.036	0.039	3.77	1.80	4.37	1.16-18.92	0.75	0.087
CT142	0.036	0.037	4.24	2.03	3.41	1.12-11.48	0.74	0.074
CT764	0.037	0.034	5.74	2.75	4.5	1.18-20.92	0.78	0.067
CT728	0.039	0.040	4.86	2.35	3.97	1.14-16.23	0.78	0.045
CT228	0.039	0.037	5.54	2.69	4.34	1.16-19.64	0.74	0.058
CT703	0.040	0.042	4.82	2.34	5.76	1.20-34.90	0.74	0.051
CT494	0.040	0.037	5.63	2.74	5.62	1.19-32.96	0.76	0.078
CT106	0.041	0.039	5.87	2.87	3.45	1.11-12.30	0.76	0.065
CT097	0.041	0.040	5.50	2.70	3.6	1.10-13.26	0.75	0.089
CT316	0.043	0.049	6.00	2.96	3.66	1.10-14.02	0.75	0.076
CT579	0.044	0.041	3.73	1.85	3.55	1.09-13.26	0.77	0.081
CT168	0.044	0.044	6.32	3.14	5.37	1.19-32.48	0.76	0.106
CT089	0.044	0.045	4.54	2.26	3.34	1.07-11.61	0.75	0.099
CT237	0.045	0.043	5.79	2.88	4.53	1.13-22.39	0.75	0.092
CT806	0.046	0.046	4.89	2.45	3.49	1.08-12.84	0.76	0.069
CT668	0.048	0.049	5.68	2.87	3.45	1.07-12.91	0.77	0.096
CT642	0.048	0.047	5.80	2.93	3.39	1.08-12.52	0.76	0.120
CT695	0.049	0.053	3.73	1.90	3.44	1.06-12.84	0.75	0.103
CT570	0.050	0.049	5.92	3.02	4.29	1.08-20.48	0.76	0.093

To examine when and how these antigens might be targeted by the host immune response during the Ct developmental cycle, expression stage predicted from Belland *et al*²¹² and predicted localisation from three bioinformatics tools were compared between the 42 differentially recognised antigens and the 441 filtered antigens (chapter 3.2.5). A χ^2 test was used to quantify over-representation or under-representation of particular expression stages or predicted localisations in the differentially recognised antigens. Late stage expressed genes were over-represented in these differentially recognised antigens, p-value 0.042 (Figure 4.6A), and both very early and very late peak expressed genes were over-represented, p-value 0.007 (Figure 4.6B). Extracellular/secreted, inner membrane and periplasmic proteins were over-represented in these antigens, p-value 0.056 (Figure 4.7).

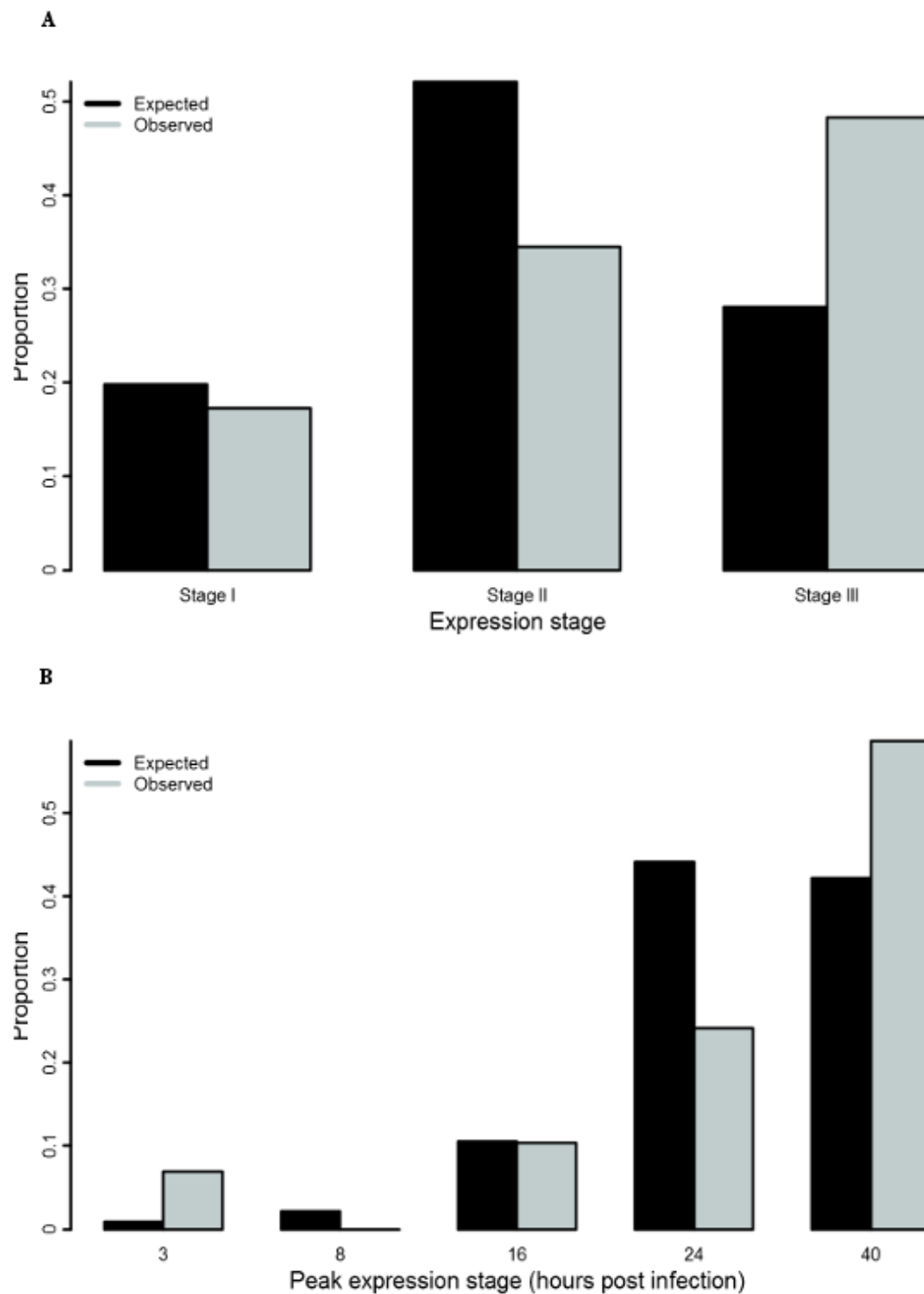


Figure 4.6. Over-representation of late and very early expressed proteins in antigens associated with susceptibility.

Proteins identified through transcriptomics as expressed late or very early in the Ct developmental cycle were over-represented in the 42 differentially recognised antigens (grey) compared with the total 441 (black).

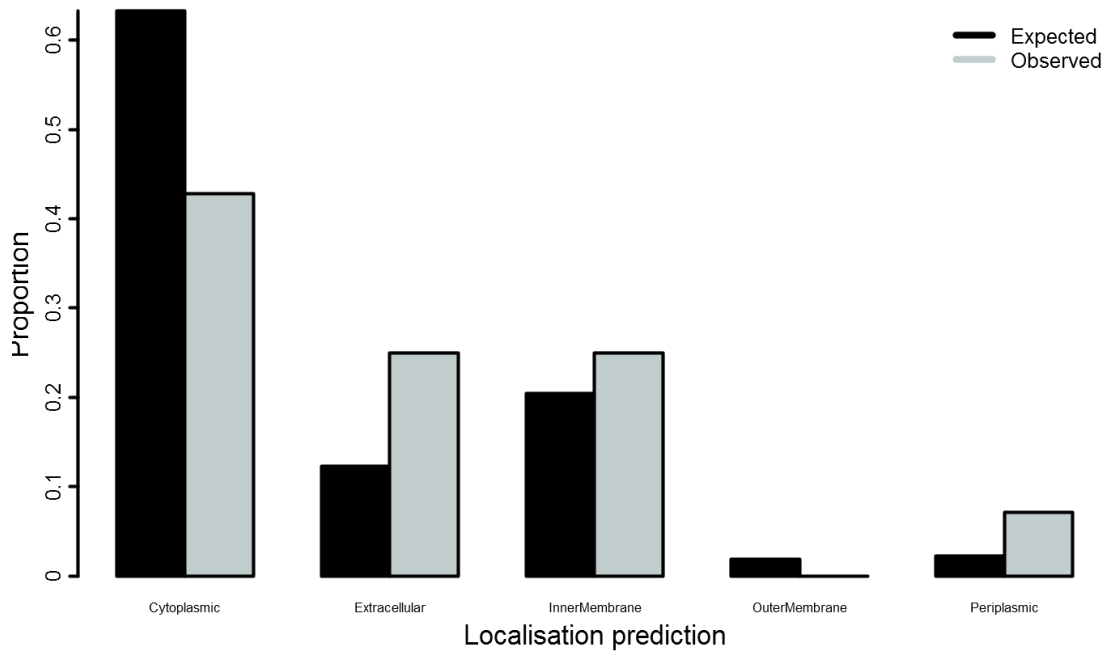


Figure 4.7. Over-representation of proteins extracellular, inner membrane and periplasm in antigens associated with susceptibility.

Proteins with a consensus localisation prediction of extracellular, inner membrane and periplasm were over-represented in the 42 differentially recognised antigens (grey) compared with the total 441 (black).

To determine the value of combining antibody responses against multiple antigens for predicting susceptibility, antigens were added step-wise to a generalised linear model adjusting for age, gender and village. Antigens were added in order, based on the strength of their association with susceptibility. Model quality (AIC) decreased as antigens were added until the 32nd antigen, the minor improvement in predictive ability (AUC) was therefore less than the complexity added for each antigen (Table 4.4). Likelihood ratio tests comparing the multivariate models with a null model including only age, gender and village had the same result. Progressive AUC improvement was likely due to overfitting as more variables were included. The median response of all 42 combined was significantly higher in unprotected individuals but was highly variable, as demonstrated by the decrease in model quality (Figure 4.8).

A model including 32 antigens was used to determine specificity and sensitivity for detecting susceptibility to infection. With 100 % specificity the maximum sensitivity was 70.00 %. With 100 % sensitivity maximum specificity was 70.00 % (Figure 4.9A).

A best compromise from this model would have 91.66 % specificity and 90.00 % sensitivity, minimising false positives without missing many true positives. To determine the value of selecting antigens based on univariate analysis, equivalent multivariate modelling was performed using antigens selected at random from the total 441. This was repeated 100 times to obtain a normal distribution. Antigens identified from univariate analysis had equivalent predictive power to those chosen at random for models of one to fifteen antigens. Models using randomly selected antigens had greater predictive power when more than fifteen antigens were included (Figure 4.9 parts B and C). Combining antigens individually associated with susceptibility to infection did not improve their predictive power.

Table 4.4. Combinatorial antibody responses increased complexity with minimal benefit to predictive power.

Generalised linear models including antigens in a step-wise manner decreased in quality until 32 antigens were included, determined by model quality (AIC) and likelihood ratio test comparing with the null model including only covariates (P-VALUE). Predictive value (AUC) showed modest progressive improvement suggestive of over-fitting.

ANTIGENS INCLUDED	AIC	AUC	P-VALUE
1	90.01	0.78	< 0.001
2	90.48	0.76	0.001
3	92.46	0.76	0.002
4	94.32	0.77	0.006
5	96.30	0.77	0.013
6	97.14	0.77	0.016
7	98.67	0.77	0.024
8	99.87	0.78	0.031
9	101.55	0.80	0.045
10	102.60	0.82	0.051
11	102.48	0.82	0.041
12	102.47	0.80	0.034
13	104.40	0.81	0.049
14	106.32	0.81	0.069
15	108.30	0.81	0.095

16	110.27	0.81	0.127
17	111.93	0.81	0.153
18	111.44	0.82	0.115
19	112.63	0.82	0.125
20	114.49	0.82	0.156
21	114.93	0.82	0.143
22	113.50	0.85	0.090
23	115.49	0.85	0.115
24	117.38	0.84	0.142
25	117.49	0.86	0.123
26	115.72	0.86	0.073
27	117.19	0.89	0.084
28	119.06	0.89	0.103
29	119.30	0.89	0.092
30	121.07	0.89	0.110
31	122.66	0.90	0.126
32	119.52	0.87	0.060
33	82.00	1.00	< 0.001

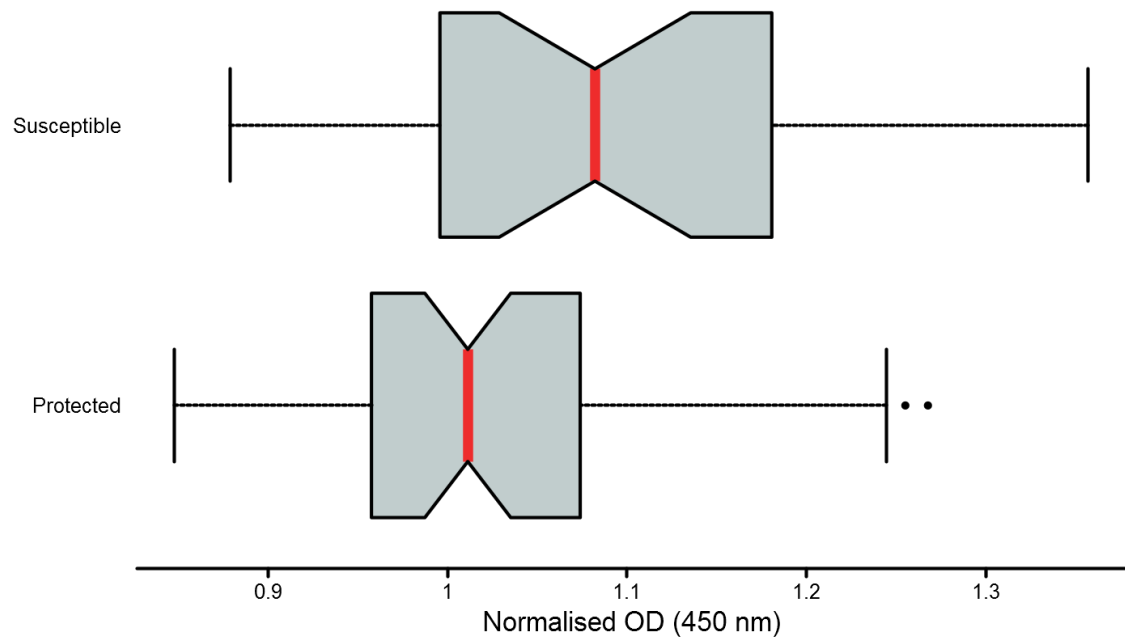


Figure 4.8. Antibody responses averaged across the top 32 differentially recognised antigens.

Average antibody responses were significantly higher in susceptible individuals but combining responses to 32 antigens significantly increased variation due to homogeneity in responses.

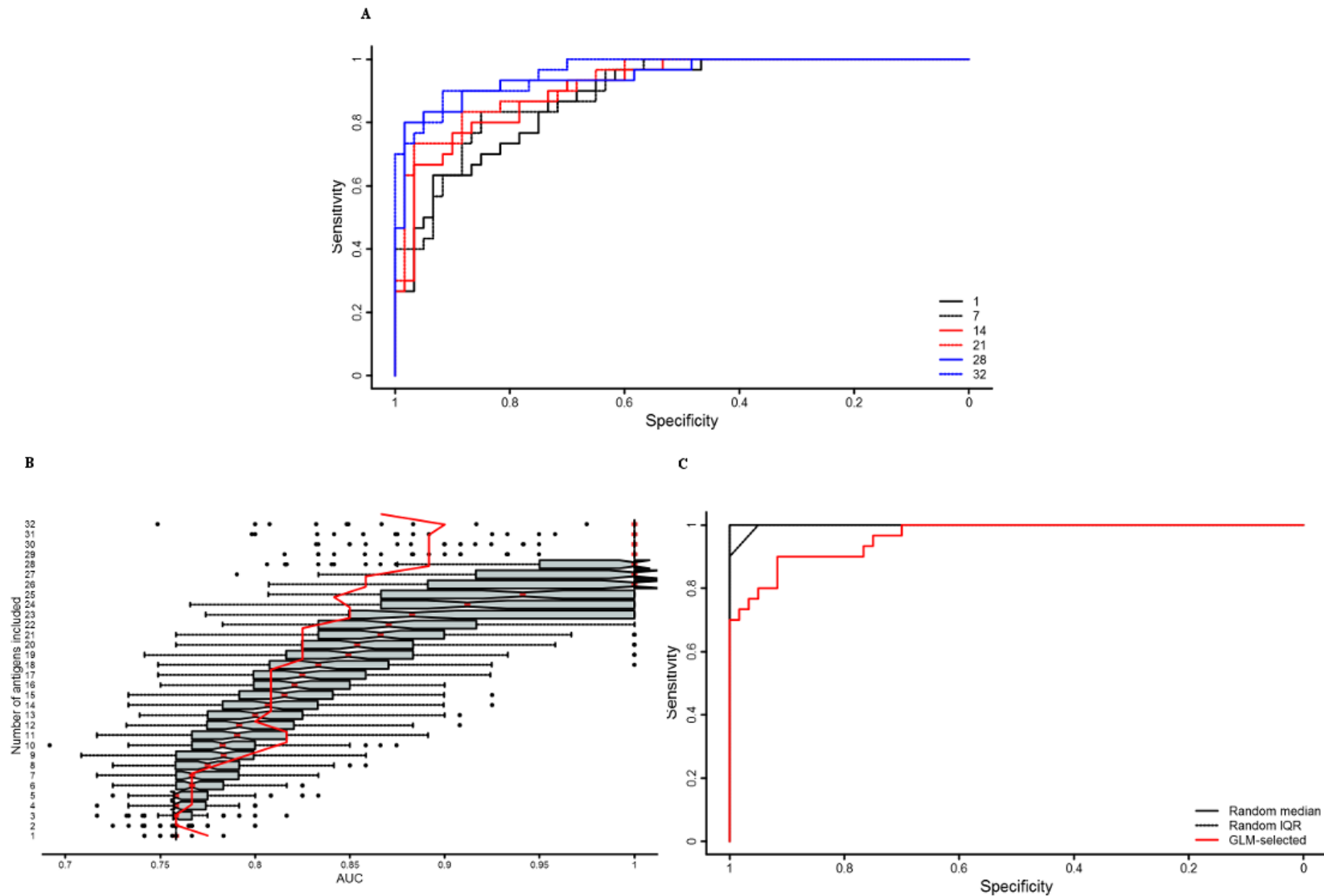


Figure 4.9. Combinatorial antibody responses did not accurately predict susceptibility.

A) Models including increasing numbers of differentially recognised antigens showed modest increases in specificity and sensitivity. B) Randomly selected antigens were as good at predicting susceptibility as antigens individually associated with lack of protection, as determined by AUC. Number of antigens included is indicated on the y-axis. C) Thirty-two randomly selected antigens outperformed the top 32 differentially recognised antigens when predicting susceptibility.

4.2.5. Combinatorial antibody responses identified by multivariate regression do not reliably predict susceptibility to infection

Combining differentially recognised antigens identified by univariate analyses did not accurately predict susceptibility. A multivariate random forests regression was performed including all 441 antigens to determine if evaluating all antigens simultaneously identified antibody targets not found in the univariate analysis that could predict susceptibility. Antigens were ranked by variable importance, defined as the mean decrease in accuracy when classifying individuals using a permutation of a variable instead of the real data. Random forests identified twelve antigens with variable importance outside 95 % of the normal distribution (Table 4.5 and Figure 4.10). Six of these were identified in the univariate analysis. Four of the remaining six were higher in susceptible individuals, CT029 and CT630 were higher in protected individuals. A multivariate model including these antigens was compared with a model of randomly selected antigens, as described previously. A multivariate model including antigens determined to be important in classification by random forests regression was not significantly better at predicting susceptibility than the average of an equivalent number of antigens selected at random (Figure 4.11).

Table 4.5. Most important antigens for classification by random forests regression.

A multivariate random forests regression was performed with all 441 antigens. These 12 antigens were in the top 2.5 % based on variable importance.

ANTIGEN	VARIABLE IMPORTANCE
CT449	0.29
CT694	0.22
CT381	0.21
CT545	0.18
CT630	0.18
CT118	0.18
CT664	0.18
CT719	0.17

CT728	0.17
CT073	0.16
CT029	0.16
CT067	0.16

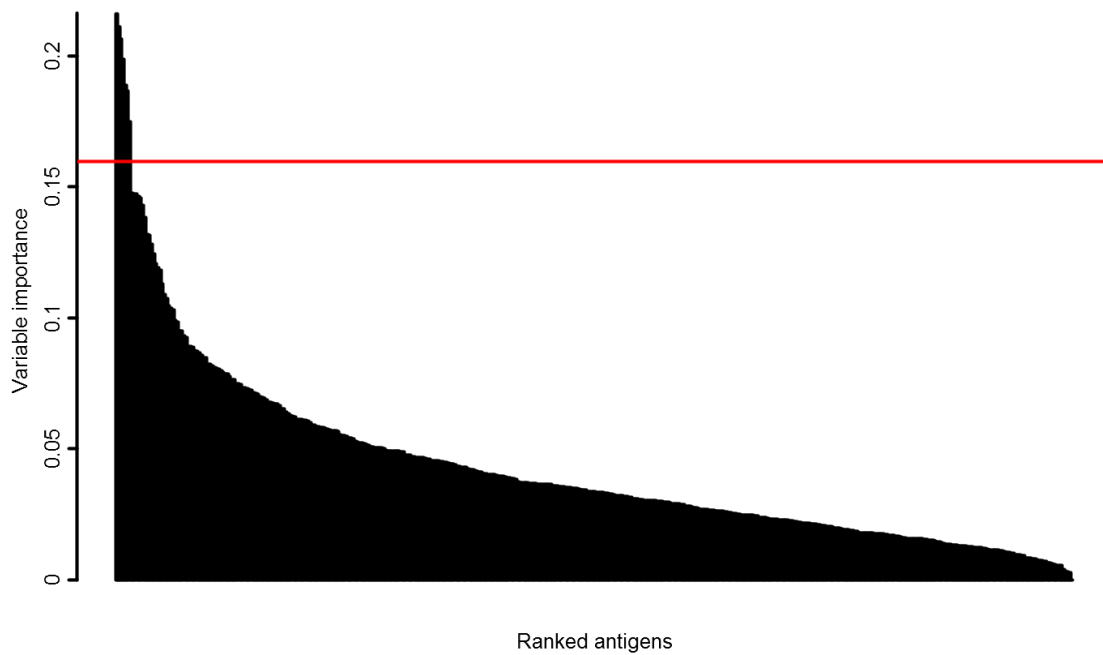


Figure 4.10. Ranked variable importance from random forests regression.

The Gini index was used to calculate variable importance. Antigens in the top 2.5 % of the distribution are indicated by the red line.

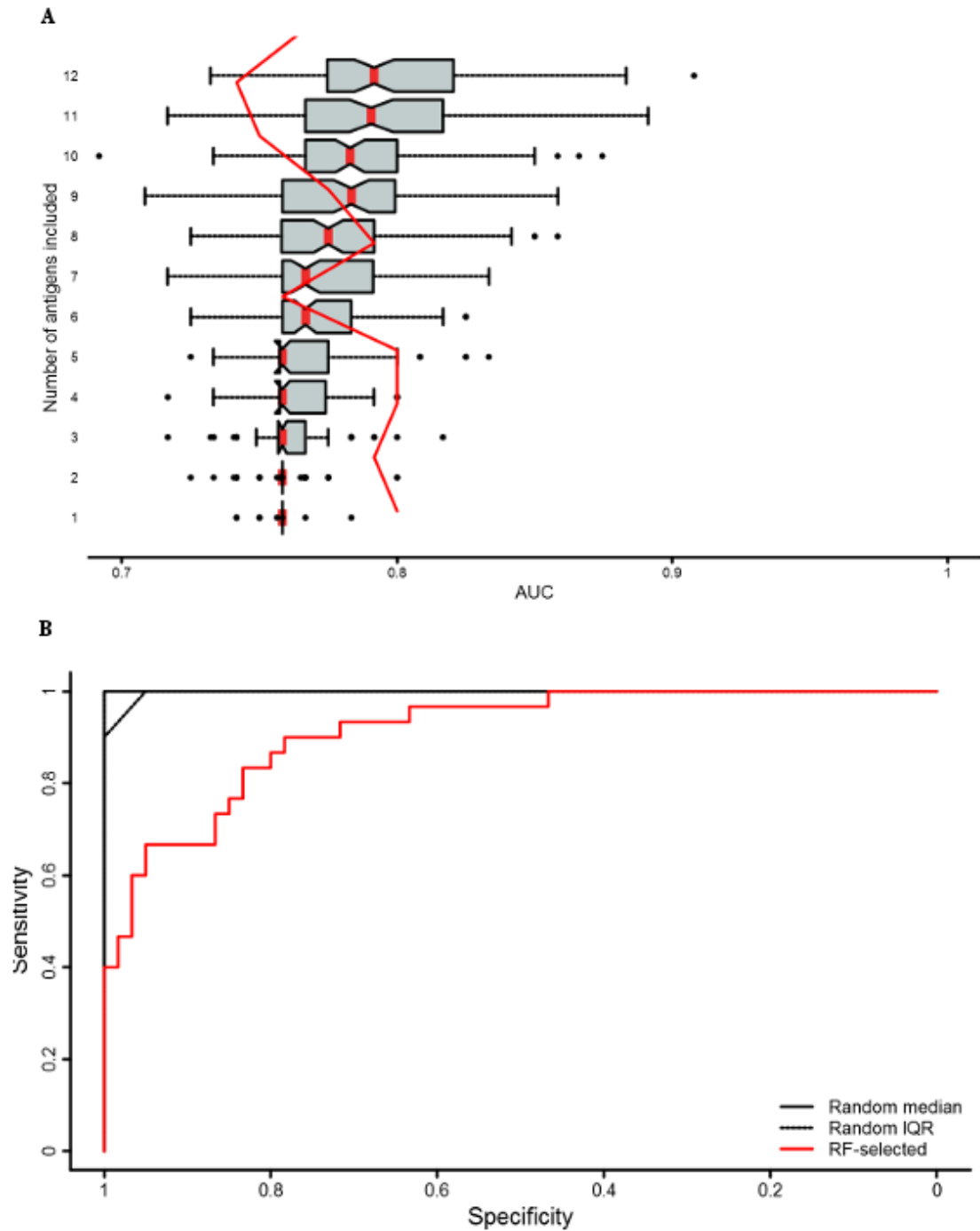


Figure 4.11. Multivariate regression did not identify antigens able to predict susceptibility.

The top 12 antigens by variable importance from a random forests regression were included in a generalised linear model. A) AUC of the model including random forests-selected antigens in a step-wise manner was no better than a model of randomly selected antigens. B) A model of the top 12 antigens was not significantly more sensitive or specific than a model of randomly selected antigens.

4.2.6. Individuals with a more focussed antibody profile are associated with protection

Approximately 10 % (42/441) of arrayed antigens included after quality filtering (Chapter 4.2.2) had significantly higher antibody responses in susceptible individuals, this boosting of a broad range of antibody responses in individuals who had repeated or long duration infections in the following six months was unexpected. The increased breadth and diversity in susceptible individuals suggested more focussed antibody responses may be protective. To investigate this, the responses of the 90 individuals were divided about the median response across all antigens to form two groups. These groups were classified as ‘globally higher’, an average response greater than the median of all antigens, and ‘globally lower, an average response lower than the median of all antigens. Univariate analysis described above was then repeated.

Twenty-six differential antibody responses were found in the group who had higher global antibody responses. Responses to each of these antigens were, all higher in susceptible individuals. Twenty-nine differential antibody responses were found in the group with lower global responses. The majority (27/29) were higher in susceptible individuals, but two were significantly higher in protected individuals (a further three were of borderline significance (Table 4.6). Of these CT029 and CT630 were both independently identified as important from the multivariate random forests regression described above.

Table 4.6. Five antigens associated with protection from infection.

Univariate analysis of all 441 antigens in individuals classified as ‘globally lower’ identified five responses associated with protection. Higher antibody responses to these antigens associated with protection from infection.

ANTIGEN	P-VALUE	P*	T	SE(T)	OR	95% CI	AUC
CT334	0.038	0.026	-12.73	6.14	0.03	0.00-0.48	0.70
CT029	0.043	0.041	-14.31	7.10	0.06	0.00-0.68	0.72
CT629	0.062	0.066	-15.87	8.51	0.09	0.01-0.89	0.69
CT391	0.065	0.063	-8.89	4.82	0.16	0.02-0.96	0.72
CT630	0.072	0.075	-17.70	9.84	0.05	0.00-0.86	0.77

The antigens identified in these three univariate analyses had surprisingly little overlap (Figure 4.12). Two antigens were found in all analyses with 40.48 %, 69.23 % and 27.59 % of targets being unique to the complete set and the two groups split around the median. CT029 was the only antigen which showed variable association with outcome. In the ‘globally higher’ group it was associated with susceptibility to infection with the opposite protection-association in the ‘globally lower’ group.

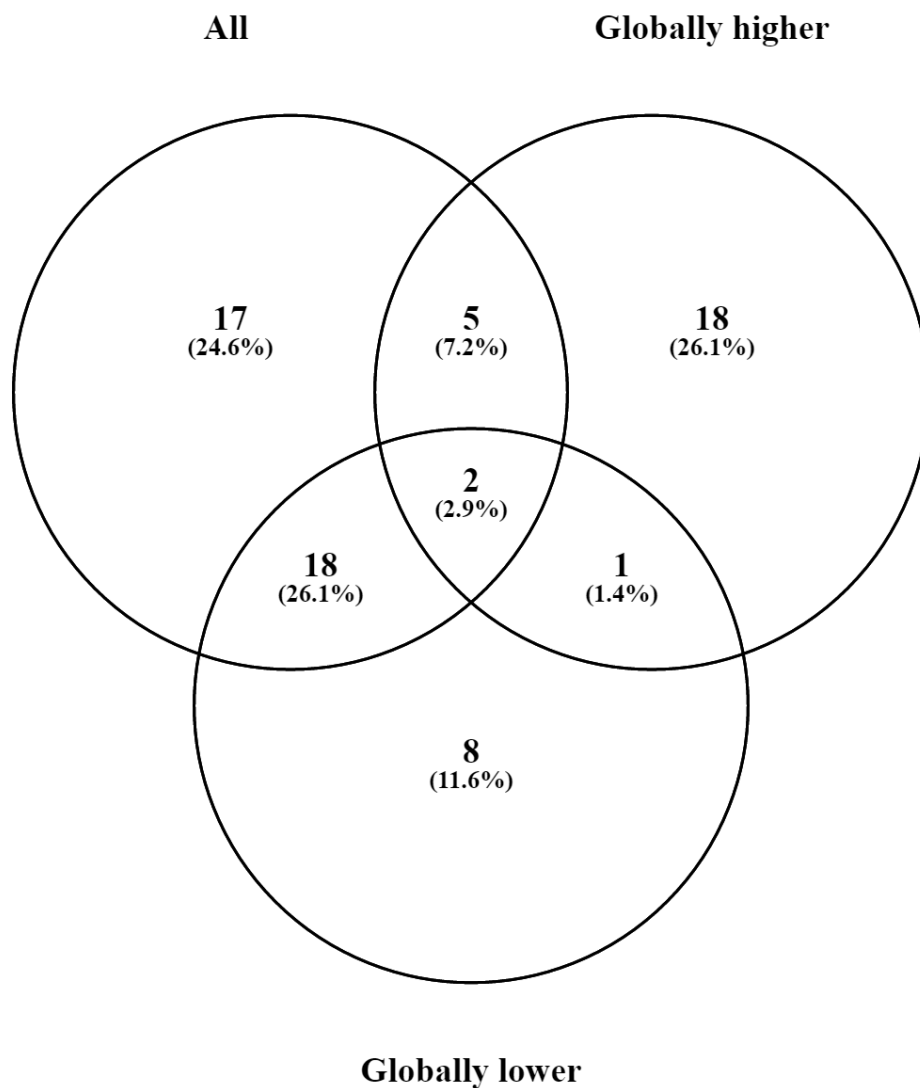


Figure 4.12. Differentially recognised antigens from the complete set, 'globally higher' individuals and 'globally lower' individuals had limited overlap.

Antigens identified in the three univariate analyses as being differentially recognised were examined for overlap. This Venn diagram was produced using Venny 2.0 (<http://bioinfogp.cnb.csic.es/tools/venny/>).

4.2.7. Selective ELISA testing of differentially recognised antigens

As described in chapter 1.4.4, follow-up work to validate results from large-scale arrays of this kind is critical. Through in-house ELISA a selection of antigens which had differential responses between protected and susceptible individuals were selected. Five of the 42 targets identified in the complete univariate analysis were inclusion membrane proteins (Incs), these are at the interface of host-Ct interactions and are plausible candidates as immune targets. Three Incs which had been previously identified as immunogenic were selected; CT813, available as a full-length GST-fusion construct, and CT118 (IncG) and CT119 (IncA), produced as peptides previously shown to be immunogenic (Prof. Bernhard Kaltenboeck personal communications, both within host-cytosol exposed regions of the proteins)²⁰⁴. Nine of the 42 targets identified were known to be secreted into the host cytosol, similarly to Incs localisation of these proteins facilitates interaction with the host immune system. Three were selected which had been identified as immunogenic previously; CT089 (CopN) and CT875, available as a full-length His-tagged proteins, and CT795, available as a full-length GST-fusion construct. Pgp3 was included as an immunodominant positive control not included on the protein microarray. Pgp3 antibody responses are purportedly related to Ct exposure and therefore should be frequently recognised in children from trachoma-endemic communities^{285, 286}.

CT089 and CT875 were produced as full-length His-tagged proteins as previously described¹⁷², provided by Prof. Steve Reed, University of Washington (Figure 4.13). CT118 (IncG) and CT119 (IncA) were synthesised as biotinylated 16-mer peptides by thinkpeptides, with purities of 92.40 % and 90.63 % respectively.

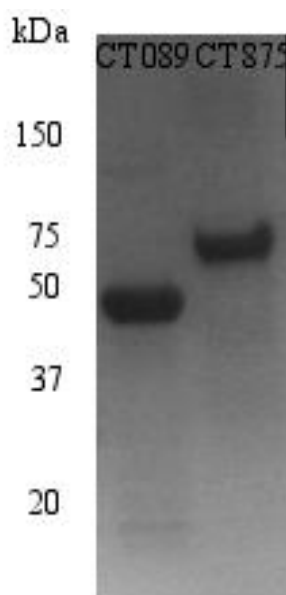


Figure 4.13: Coomassie blue stained SDS-PAGE gel of CT089 and CT875.

The molecular weight markers on the left are in kDa.

CT795 and CT813 were both available as GST-fusion constructs. Due to high levels of non-specific binding from the anti-GST antibody, even when used 10-fold more dilute than recommended, bands of interest were highlighted in red when there was no clear dominant band. Expression of both CT795 and CT813 was inducible in initial trials at 37 °C (Figure 4.14A). CT795 aggregated and CT813 aggregated or was found in insoluble inclusion bodies using the recommended conditions for induction (Figure 4.14 parts B and C); 30 °C for 3 hours with 0.2 mM IPTG. This was unsurprising for the membrane-localised CT813 but not for CT795 which is secreted by Ct.

These proteins were co-expressed with molecular chaperones to try and improve expression and solubility^{202, 203}. These chaperones are expressed from plasmids previously transformed into competent *E. coli* prior to transformation with the GST-fusion constructs. CT795 was still commonly aggregated however some protein was also soluble using the pKJE7 (Figure 4.15A lanes 10-12) or the pTf16 (Figure 4.15A lanes 7-9) plasmids, expression levels were significantly higher using the latter. CT813 was no longer within inclusion bodies, with both plasmids it was equal parts aggregated and soluble (Figure 4.15A lanes 1-6). CT795 was able to be purified using glutathione beads from the soluble fraction after co-expression with the pTf16 plasmid, CT813 was able to be purified from the soluble fraction after co-expression with either plasmid,

expression levels were marginally higher with pKJE7 (Figure 4.15B). Subsequently the GST-fusion was cleaved during purification to try and minimise the impact of non-specific binding and reactivity in subsequent serological work. CT795 was co-expressed with pTf16 and CT813 was co-expressed with pKJE7, they were successfully purified with cleavage of the GST-fusion however there were still many non-specific bands present (Figure 4.15C). These bands were likely a mixture of bacterial contaminants and the molecular chaperones, size exclusion chromatography was performed using a Superdex 200 column with an Äkta Purifier to try and separate out the Ct proteins. There was a small reduction in contaminants however the chaperones were still present post-purification, due the abundance of the chaperones after purification and their unknown immunogenicity they were deemed unsuitable for the following serology. Time-constraints and the availability of other proteins from the array meant no further attempts were made to purify CT795 and CT813.

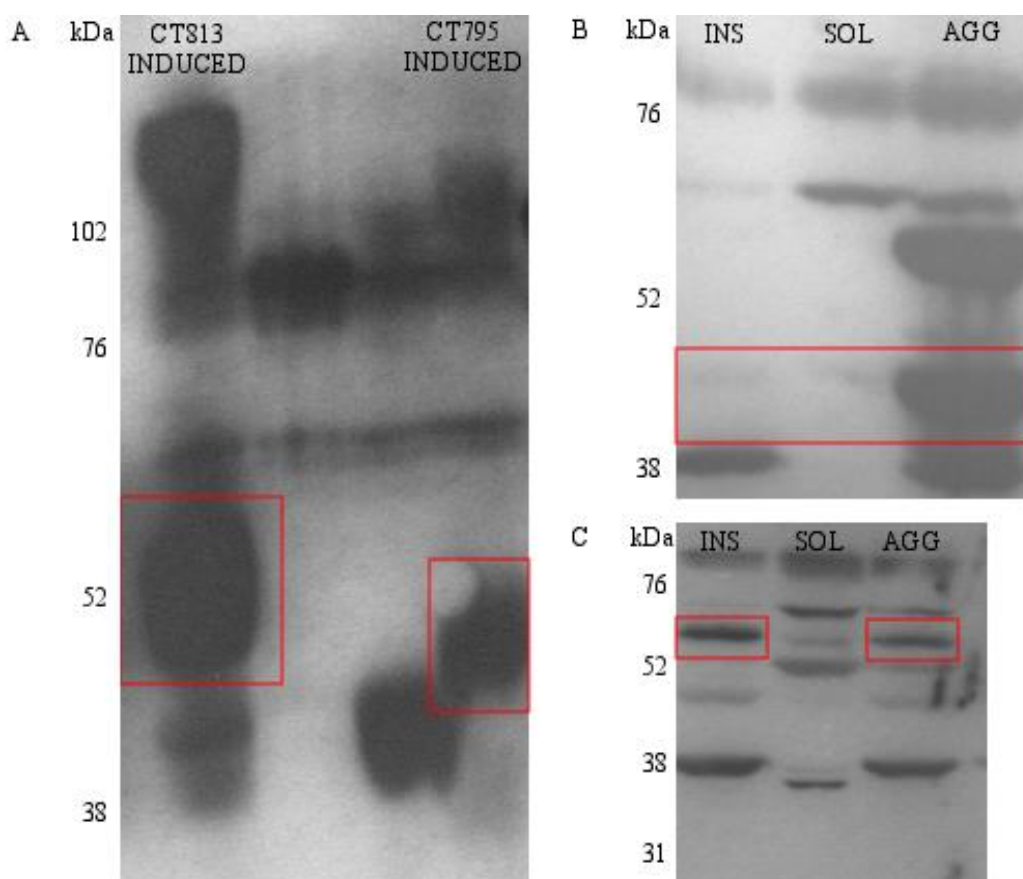


Figure 4.14. Expression trials of CT795-GST and CT813-GST.

Western blots were incubated with an anti-GST monoclonal antibody to bind the GST-fusion moiety. The lysates were separated by ultra-centrifugation into aggregate (AGG), soluble (SOL) and insoluble fractions (INS). The relevant bands are highlighted in red. A) Successful expression of CT795-GST and CT813-GST after IPTG induction in *E. coli* lysates. B) CT795-GST was primarily aggregated when induced at 30 °C for 3 hours with 0.2 mM IPTG. C) CT813-GST was aggregated or insoluble when induced under the same conditions.

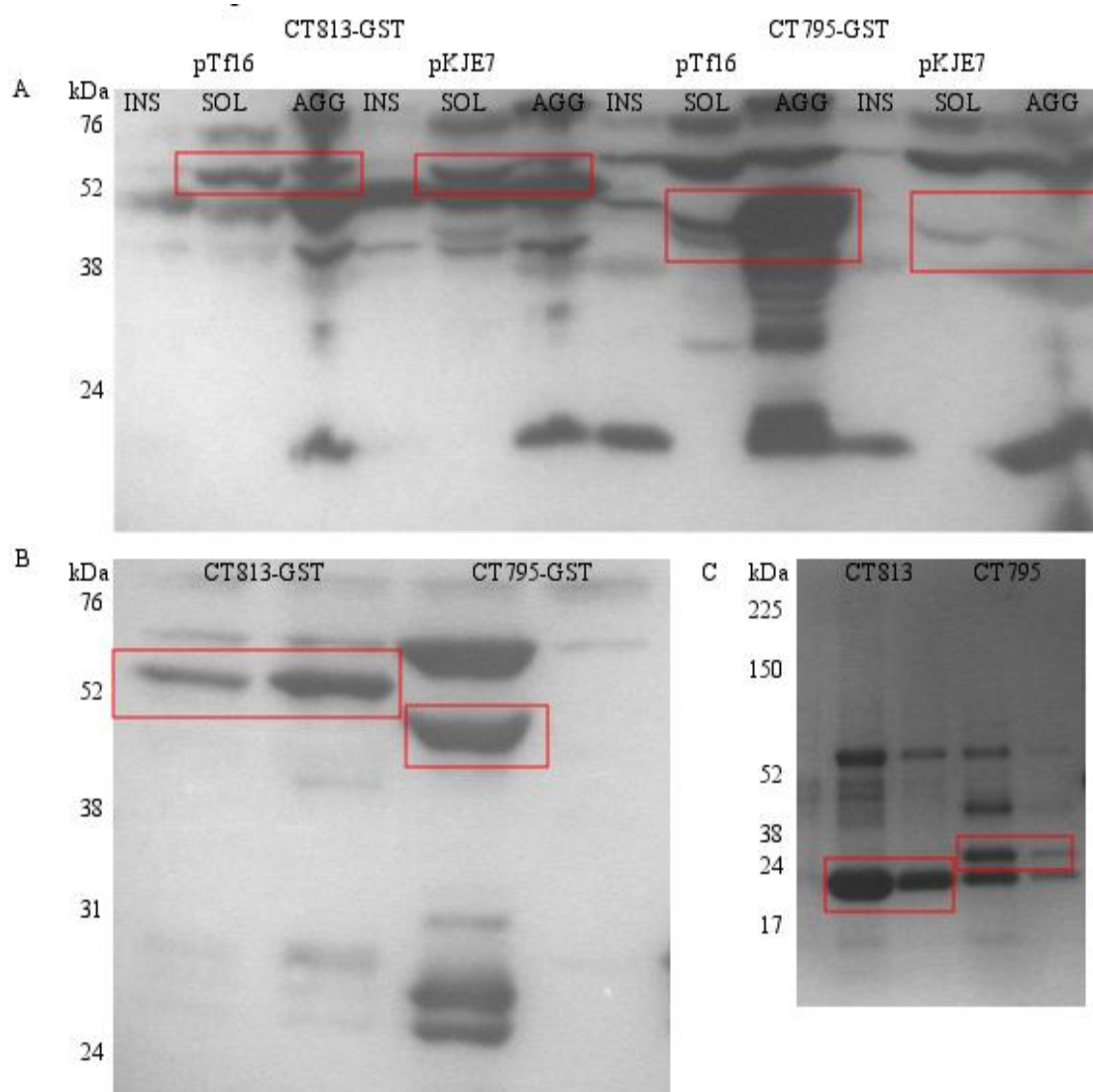


Figure 4.15. Co-expression trials of CT795-GST and CT813-GST with chaperone plasmids.

Western blots were incubated with an anti-GST monoclonal antibody to bind the GST-fusion moiety. The lysates were separated by ultra-centrifugation into aggregate (AGG), soluble (SOL) and insoluble fractions (INS). The relevant bands are highlighted in red. A) Expression was trialled using the pKJE7 and pTf16 chaperone-expressing plasmids. Both proteins were partially soluble with either chaperone set, CT795-GST expression was significantly greater with pTf16. B) Both proteins were successfully purified using glutathione beads, non-highlighted dominant bands indicated co-purification of the molecular chaperones. C) Both proteins were successfully purified after cleavage of the GST-fusion moiety, the chaperones were still present at high levels.

Pgp3 was soluble using the recommended induction conditions described above both with and without cleavage of the GST-fusion (Figure 4.16 parts A and B). Pgp3 was purified further by size-exclusion chromatography. No non-specific bands were detectable by silver staining after this purification (Figure 4.16C). When compared with the original purification of Pgp3 little difference was detectable using Coomassie-blue staining (Figure 4.16D), however there was a clear difference in purity when examined by silver staining (Figure 4.16E).

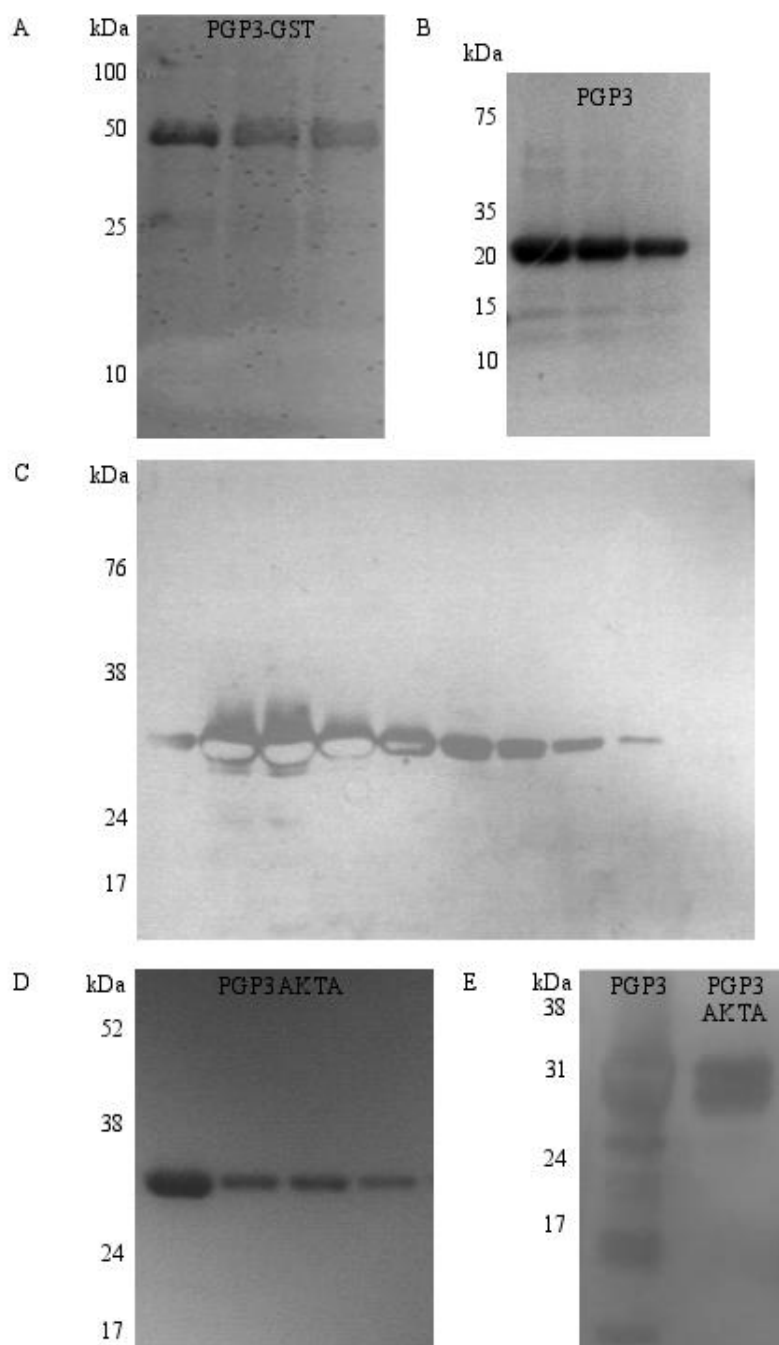


Figure 4.16. Expression and purification of Pgp3

A) Coomassie-blue stained SDS-PAGE gel of Pgp3-GST was expressed in a soluble form when induced at 37 °C for 3 hours with 0.2 mM IPTG and was purified using glutathione beads. B) Coomassie-blue stained gel Pgp3 was purified after cleavage of the GST-fusion moiety. C) Impurities were removed from Pgp3 by size exclusion-based chromatography. No non-specific bands were detectable by silver staining. D) Pgp3 pre (left) and post-AKTA purification had similar purity was Coomassie-blue staining. E) Silver staining showed greater purity of Pgp3 after further purification by size exclusion-based chromatography.

4.2.8. ELISA validation showed variable association with a lack of protection

The initial ELISA protocol was adapted from methods used previously in trachoma or urogenital Ct infection^{140, 161}. Randomly selected sera were used from three different study populations to optimise conditions with a positive control Pgp3 (chapter 3.4). A serum dilution of 1/500 combined with a Pgp3 concentration of 1 µg/ml produced the greatest breadth of responses (Figure 4.17 parts A and B). When tested with these conditions a secondary antibody dilution of 1/ 30000 had the strongest discriminatory power (Figure 4.17C). Responses against GST-cleaved Pgp3 and Pgp3-GST had minimal difference in absorbance when tested on a small set of sera.

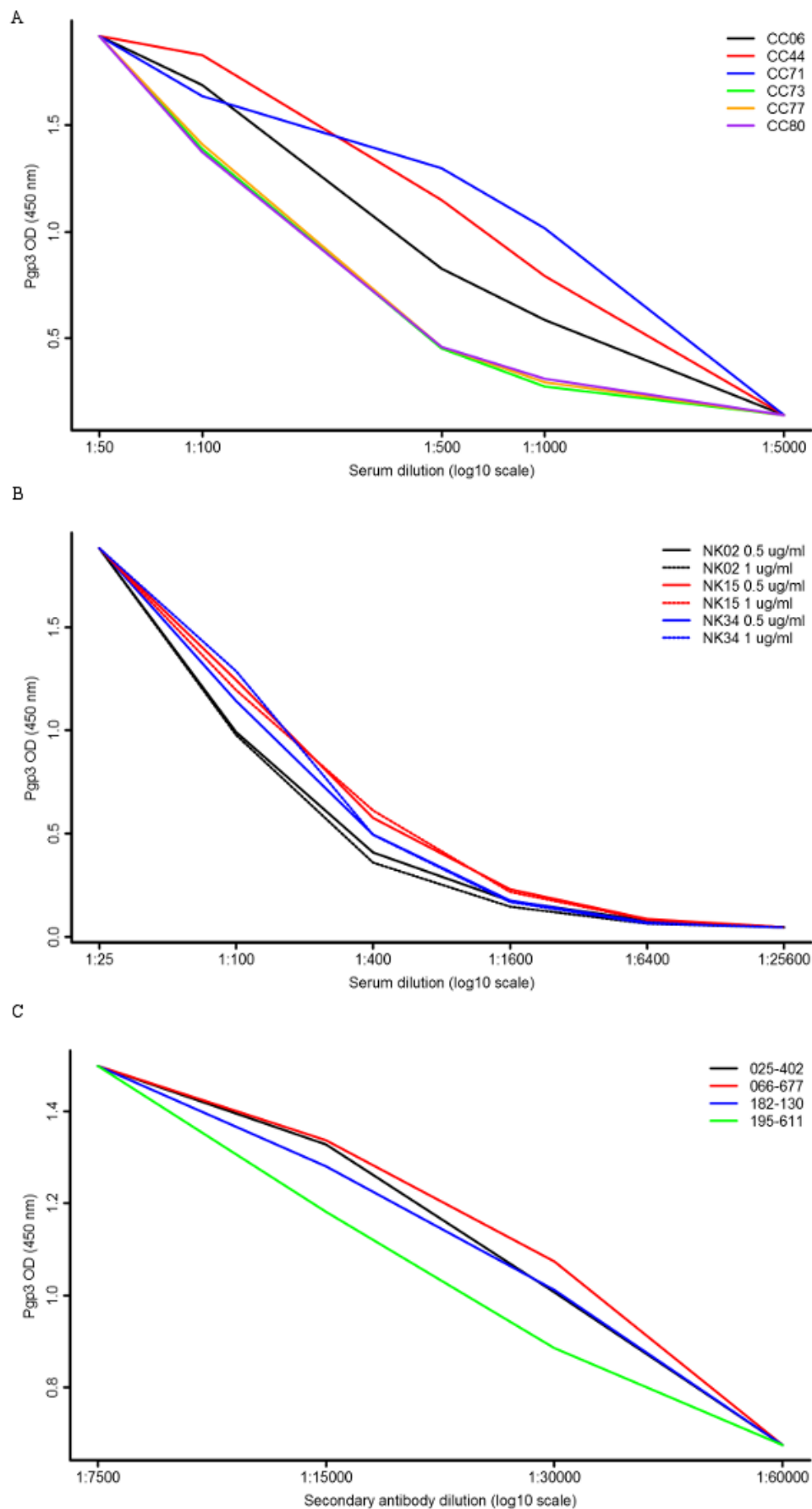


Figure 4.17. Optimisation of an in-house ELISA protocol.

All optimisation was performed with Pgp3. A) A serum dilution of 1/500 produced the greatest breadth of responses. B) Pgp3 concentration of 1 μ g/ml (dashed line) produced the greatest breadth of responses. C) A secondary antibody dilution of 1/30000 provided the greatest breadth without a high background response.

The ELISA protocol was validated by visually inspecting how Pgp3 responses changed with age and by comparing results with two published Pgp3-based ELISA protocols (chapter 3.5). The median Pgp3 response increased significantly from children (one to nine years) to teenagers and young adults (ten to nineteen years) and from this group to adults (twenty years and above) in a trachoma-endemic community in The Gambia, p-values of 0.003 and 0.006 respectively (Figure 4.18A). The ELISA results were not strongly correlated with those from a Pgp3-ELISA developed independently by Dr Myra McClure's group at Imperial College London²³⁴, rho value 0.14 and a p-value of 0.440. (Figure 4.18B). In this small set of sera, samples with strong Pgp3 responses were correlated. Correlation was weaker in samples with Pgp3 responses less than 0.5 OD, these responses were mostly non-specific background absorbance. The Pgp3 ELISA had very strong correlation with a Pgp3-ELISA developed independently by Dr Diana Martin at Center for Disease Control (CDC), Atlanta²⁸⁶, rho value 0.88 and a p-value < 0.001 (Figure 4.18C). This was tested on a significantly larger set of samples with a greater range of Pgp3 responses. The Pgp3-ELISA methodology developed for this study performed comparably with these published tests.

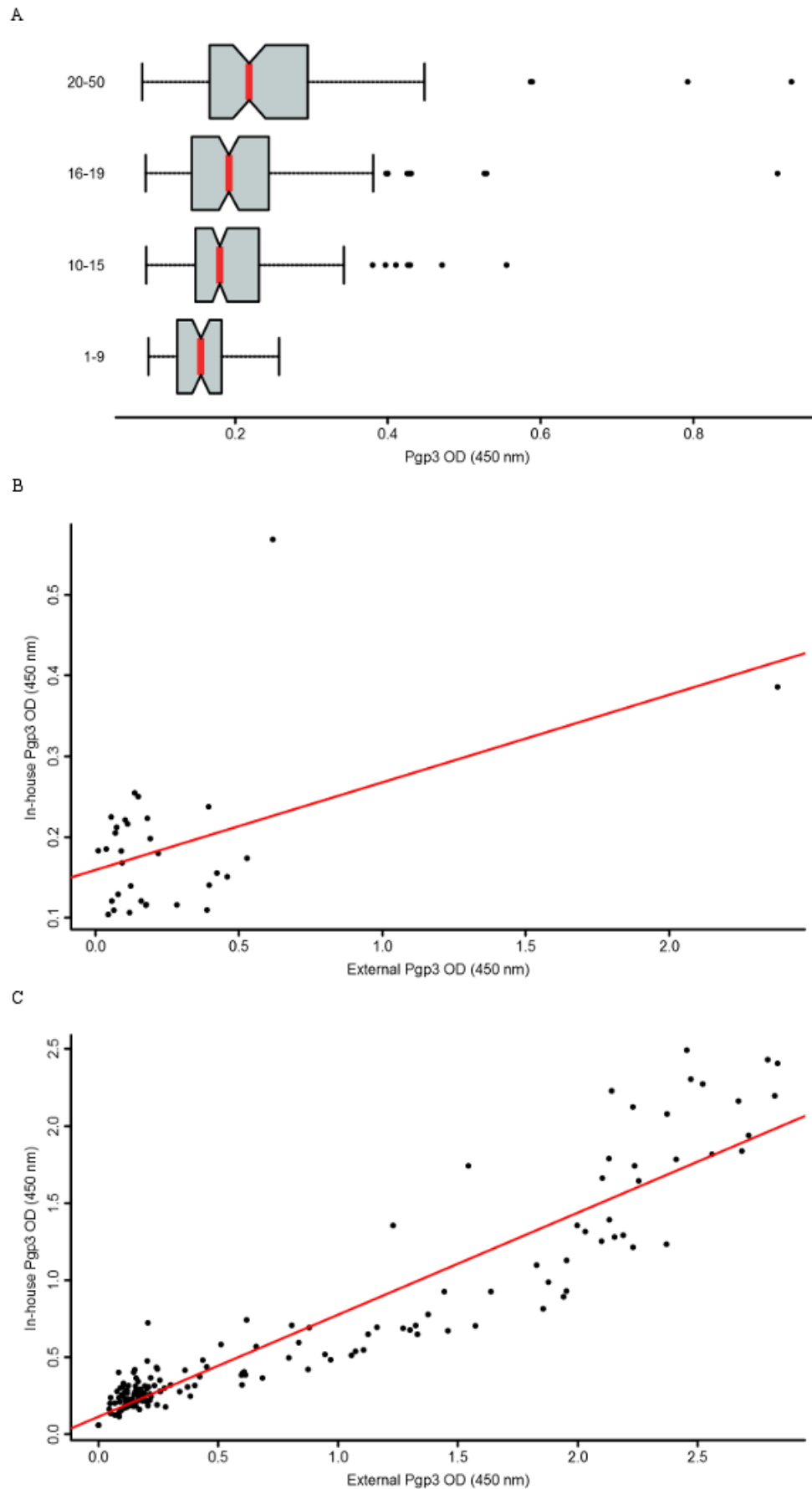


Figure 4.18. Validation of an in-house Pgp3 ELISA.

The ELISA methodology was validated by; A) examining if Pgp3 response increased with age and B) and C) comparing results with two published Pgp3-ELISA protocols. A linear model of ELISA results from the in-house and the two published methods was used to fit the line (red). 120

For the protein microarray validation, 90 serum samples that were previously tested by microarray were retested by ELISA. A further 40 samples were included from the baseline of the cohort sample that had insufficient follow-up data for inclusion in the longitudinal analysis. The total 130 sera had a similar make-up in terms of age, gender and village of participants (Table 4.7). These sera were tested for antibody responses to CT089, CT875, IncG, IncA and Pgp3.

Table 4.7. Patient demographics in protected and susceptible groups, complete sera set.

Associations with susceptibility to infection were determined using a generalised linear model.

	PROTECTED (IMMUNE)	SUSCEPTIBLE (NON-IMMUNE)	P-VALUE
NUMBER	95	35	NA
AGE IN YEARS (95% CI)	9.00 (2.00-14.00)	9.00 (1.85-13.15)	0.982
FEMALE (N [%])	42 (44.21)	13 (37.14)	0.470
VILLAGE (N [%])			0.002
- 1	13 (13.68)	18 (51.43)	
- 2	10 (10.53)	2 (5.71)	
- 3	1 (1.05)	0 (0.00)	
- 4	14 (14.74)	3 (8.57)	
- 6	14 (14.74)	5 (14.29)	
- 7	34 (35.79)	6 (17.14)	
- 8	9 (9.47)	1 (2.86)	

ELISA responses for CT089 and CT875 were poorly correlated with responses from the array, rho values of 0.14 and 0.19 and p-values of 0.203 and 0.014 respectively (Figure 4.19A and 4.20A). ELISA responses for both IncG and IncA had almost no correlation with responses from the array, rho values of 0.01 and < -0.01 and p-values of 0.948 and 0.972 respectively (Figures 4.21A and 4.22A). Despite this, responses to all four were still significantly higher in unprotected individuals, p-values of 0.003, 0.015, 0.005 and 0.008 respectively (Figures 4.19B, 4.20B, 4.21B and 4.22B). The positive control Pgp3 was also significantly higher in unprotected individuals, p-value < 0.001 (Figure 4.23).

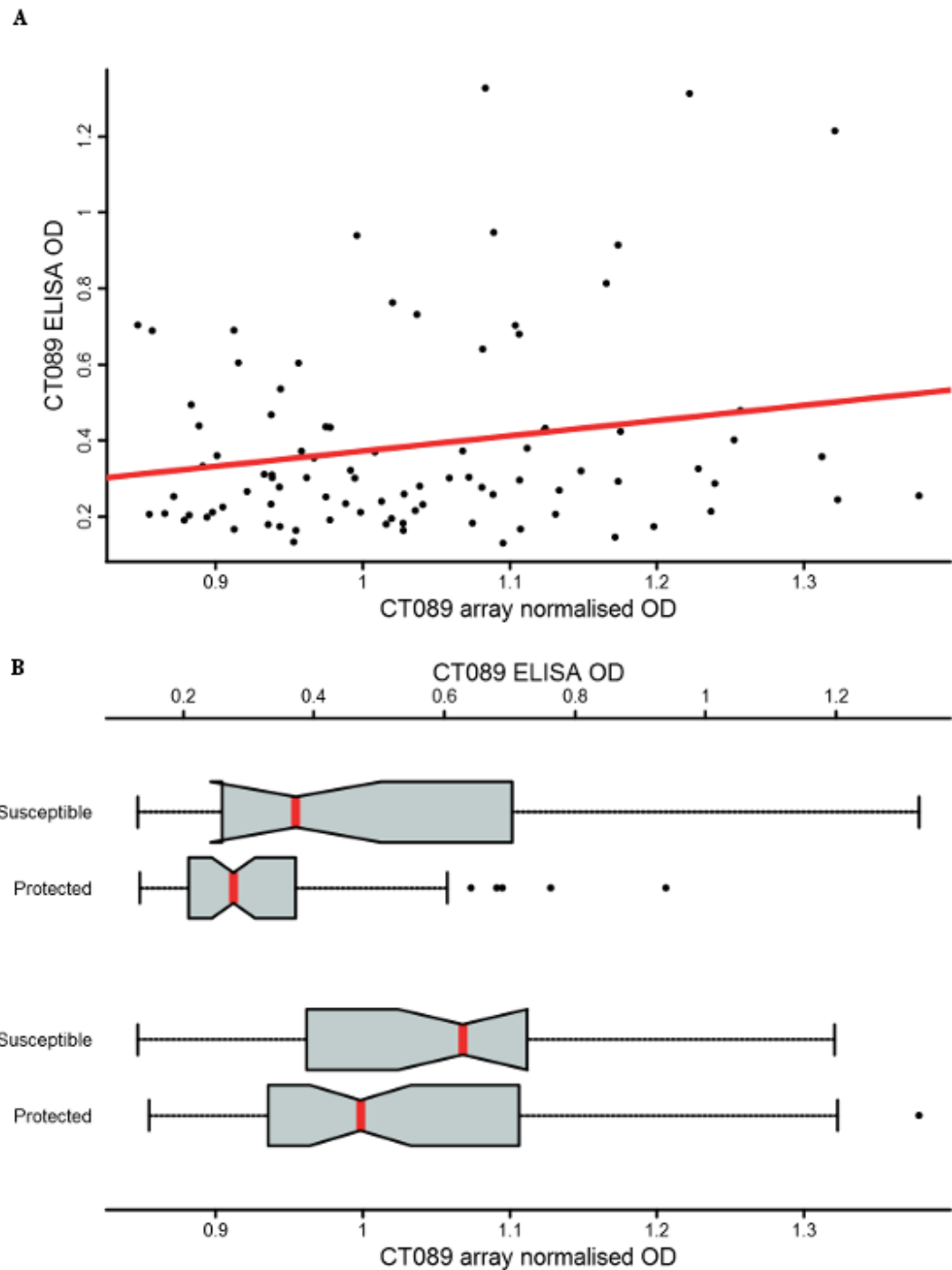


Figure 4.19. CT089 correlation between ELISA and array results.

Responses to CT089 were retested in 90 arrayed serum using and in-house ELISA. A) Correlation between ELISA and array results was variable. B) Higher responses to CT089 remained associated with susceptibility to infection. A linear model of results from the ELISA and array was used to fit the line (red).

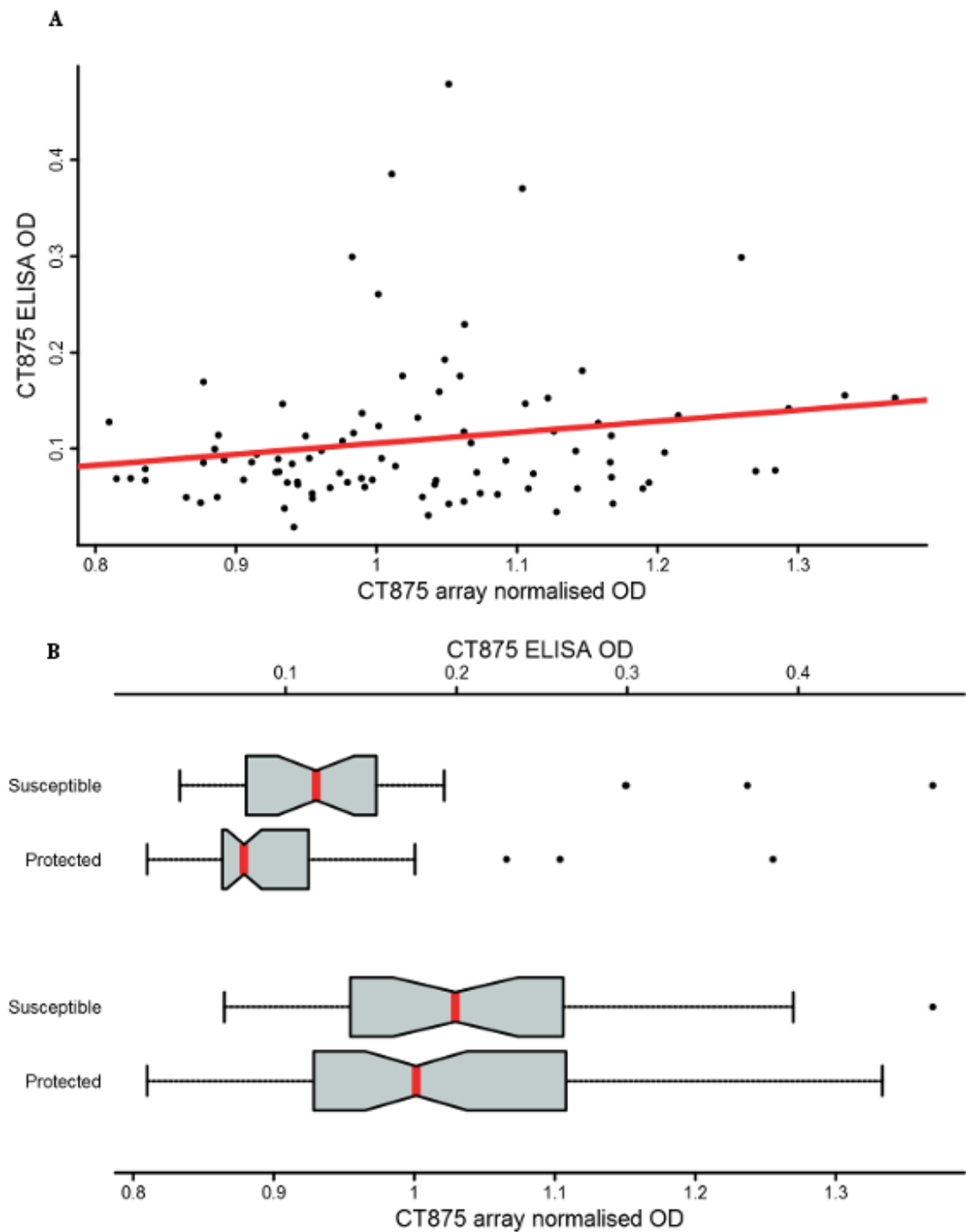


Figure 4.20. CT875 correlation between ELISA and array results.

Responses to CT875 were retested in 90 arrayed serum using and in-house ELISA. A) Correlation between ELISA and array results was poor. B) Higher responses to CT875 remained associated with susceptibility to infection. A linear model of results from the ELISA and array was used to fit the line (red).

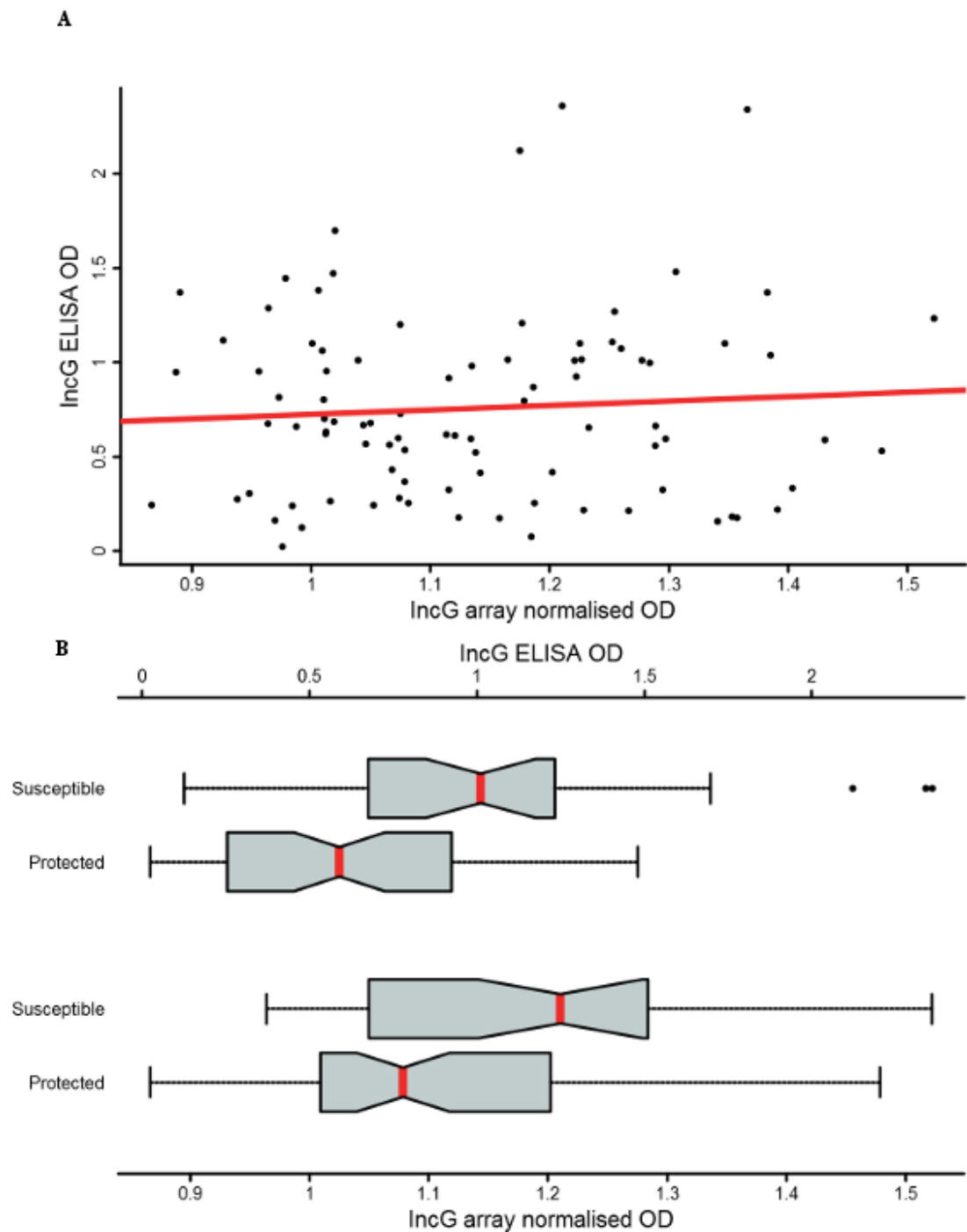


Figure 4.21. IncG correlation between ELISA and array results.

Responses to IncG were retested in 90 arrayed serum using and in-house ELISA. A) There was no correlation between ELISA and array results for IncG. B) Higher responses to IncG remained associated with susceptibility to infection. A linear model of results from the ELISA and array was used to fit the line (red).

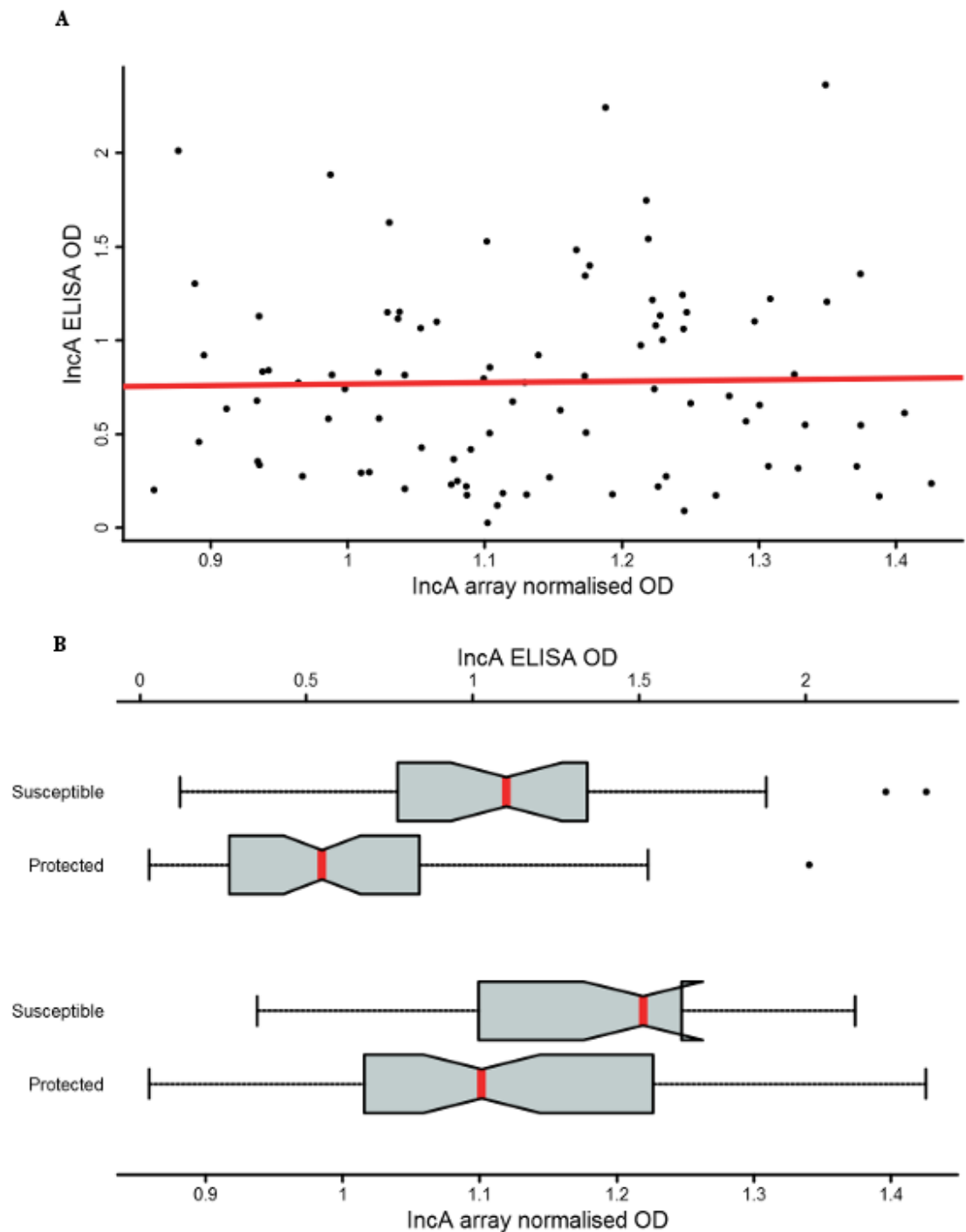


Figure 4.22. IncA correlation between ELISA and array results.

Responses to IncA were retested in 90 arrayed serum using and in-house ELISA. A) There was no correlation between ELISA and array results for IncA. B) Higher responses to IncA remained associated with susceptibility to infection. A linear model of results from the ELISA and array was used to fit the line (red).

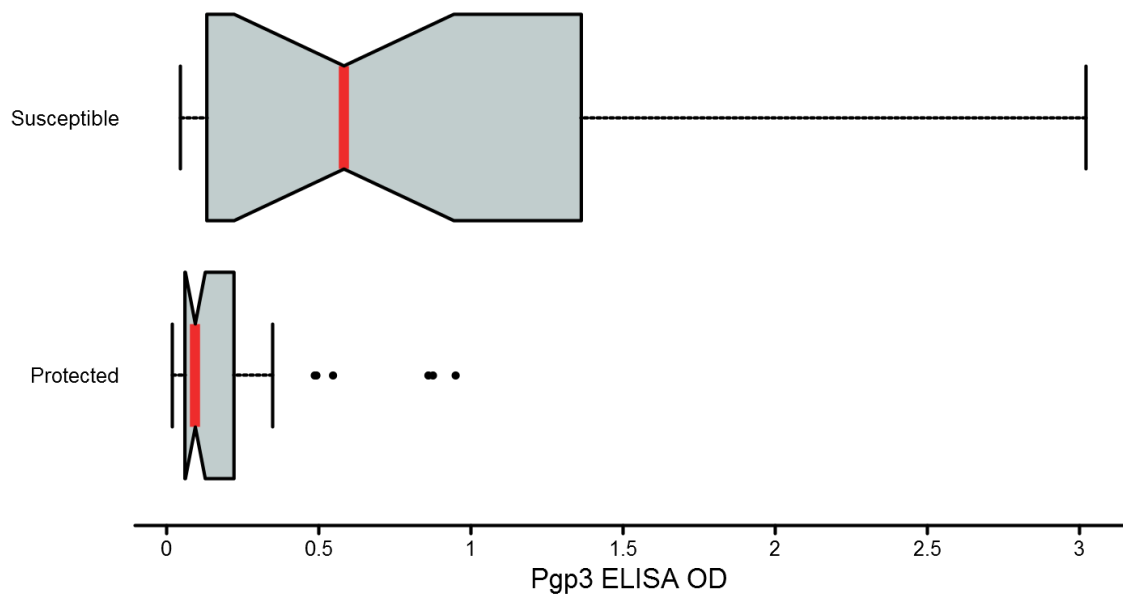


Figure 4.23. Pgp3 responses in 90 arrayed sera.

Responses to Pgp3 were tested in 90 arrayed serum using an in-house ELISA. Higher Pgp3 responses were significantly associated with susceptibility to infection.

Discrepancy between ELISA responses and those from the array were investigated using CT089 and CT875 as examples. In both cases but most notably for CT089 there were two populations of responses, one was concordant between the ELISA and the array the other was discordant. These were approximately split around the line of best fit (Figure 4.19A). The discordant results were all low by ELISA but varied from low to high on the array, meaning they were caused by either poor ELISA sensitivity or poor array specificity. Those deemed discordant by CT089 and CT875 were compared and classified as; “concordant” where responses between ELISA and the array were correlated, “both discordant” where responses were not correlated for either antigen or “one discordant” where responses to only one antigen were not correlated between the tests (Table 4.8 and Appendix Figure 2).

Table 4.8. Agreement of concordant and discordant sera from CT089 and CT875.

CT089 ELISA VS ARRAY			
CT874 ELISA VS ARRAY		CONCORDANT	DISCORDANT
	CONCORDANT	19	8
	DISCORDANT	12	51

ELISA responses to all five antigens were compared between these groups. Responses within the “both discordant” group were significantly lower than both other groups for all antigens (Figure 4.24). This suggested a problem with these sera rather than the array, potentially due to differential storage with/without glycerol since collection. A three-fold cross-validation of the generalised linear model described earlier was used to test if these discordant sera were producing false associations, the data were split to maintain the 2:3 ratio of protected to unprotected in the training and validation sets. Eighty-eight percent (37/42) of the antigens identified as differentially recognised previously were significantly associated with a lack of protection, combined p-values ≤ 0.05 . This means the lack of concordance was due to sera upon retesting by ELISA not recognition or reactivity on the protein microarray. Excluding these discordant sera, correlation was improved for all four arrayed antigens. Responses remained higher in susceptible individuals and this was significant for CT089 and CT875 (Table 4.9).

Table 4.9. ELISA and array results after excluding “both discordant” sera.

Associations with susceptibility to infection were retested. The association was calculated excluding the sera classified as “both discordant”. The odds ratio and confidence intervals were calculated from the array results.

ANTIGEN	SPEARMAN'S RHO	ELISA P-VALUE	GLM P-VALUE	OR (95 % CI)
CT089	0.62	< 0.001	0.022	56.53 (2.82-3486.86)
CT875	0.67	< 0.001	0.027	4.75e ¹³

				(3.74e ⁴ -3.07e ²⁸)
INCG	0.27	0.094	0.078	5.81 (0.92-52.13)
INCA	0.32	0.042	0.191	3.60 (0.55-2.86)

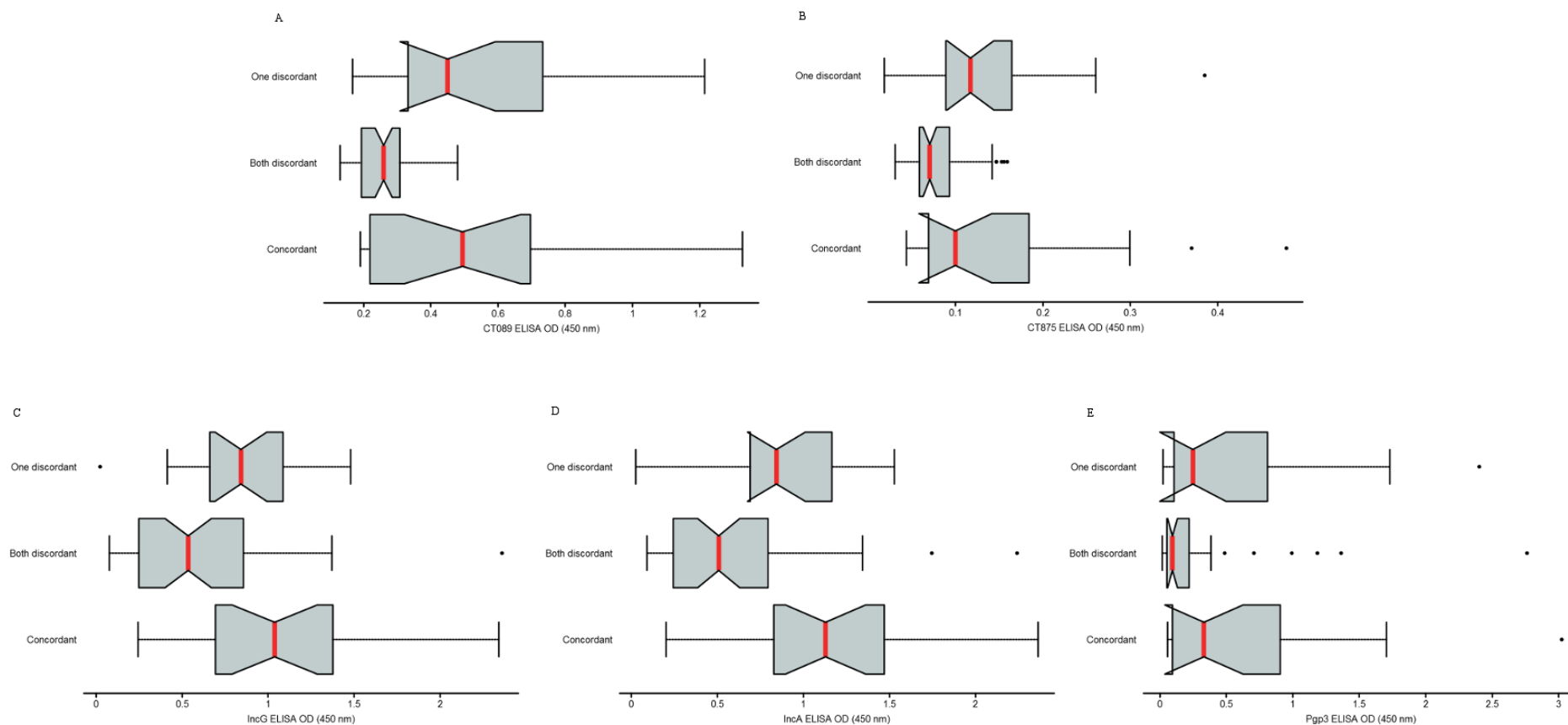


Figure 4.24. ELISA responses in “both discordant” sera were lower for all antigens.

ELISA results for each antigen were grouped using the definitions from Table 4.8. Responses were lower in “both discordant” for all tested antigens; A) CT089, B) CT875, C) IncG, D) IncA and E) Pgp3.

To further test the association of these antigens with susceptibility to infection, we tested the complete 130 samples from the longitudinal cohort both with and without the discordant sera described previously (Table 4.10). Associations were consistent with results from the array for all four antigens.

Table 4.10. ELISA results from the complete 130 sera.

The four arrayed antigens were tested on the complete set of 130 sera (COHORT) and excluding “both discordant” sera (COHORT*). Higher responses to all antigens remained significantly associated with susceptibility to infection.

ANTIGEN	SERA	PROTECTED MEDIAN (IQR)	SUSCEPTIBLE MEDIAN (IQR)	P- VALUE	OR (95% CI)
CT089	COHORT	0.29 (0.22-0.41)	0.36 (0.27-0.70)	0.027	10.47 (1.47-101.25)
	COHORT*	0.33 (0.24-0.46)	0.60 (0.33-0.81)	0.007	37.35 (3.42-709.96)
CT875	COHORT	0.09 (0.07-0.11)	0.12 (0.08-0.15)	0.008	4.97e ⁴ (36.78-3.10e ⁸)
	COHORT*	0.10 (0.08-0.12)	0.13 (0.11-0.18)	0.004	3.43e ¹⁰ (2.53e ⁴ -5.00e ¹⁸)
INCG	COHORT	0.37 (0.22-0.69)	0.96 (0.58-1.16)	0.002	5.85 (2.09-19.25)
	COHORT*	0.37 (0.23-0.79)	0.96 (0.60-1.21)	0.026	4.40 (1.27-18.03)
INCA	COHORT	0.35 (0.22-0.68)	0.97 (0.68-1.23)	0.002	5.15 (1.90-15.74)
	COHORT*	0.35 (0.22-0.83)	0.97 (0.70-1.24)	0.042	3.79 (1.10-14.95)

4.3. Discussion

In this study ocular Ct infection data were used from a six-month longitudinal cohort in Gambian children from a trachoma-endemic community. To attempt to identify correlates of immunity, individuals who were able to successfully resolve infection were compared with those who were susceptible to repeated or long durations infection over the following six months. From screening the Ct protein microarray for antibody responses in this cohort the relationship between individual serological immune responses and the variable acquisition and resolution of ocular Ct infection was characterised. Antibody responses were significantly more focussed in children who were able to resolve infection while heightened responses to 42 antigens were associated with susceptibility to infection and longer durations of infection. By focussing on children with less diverse global antibody profiles, 6 antigens were identified where heightened responses were associated with ability to resolve infection. Independent validation of these antigens associated with long-term Ct infection outcome through further serological testing was variably successful, highlighting the need for substantiation of findings from large-scale array-based techniques as described here.

4.3.1. Longitudinal evidence of Ct infection reflects differences in immunity

Studies of ocular Ct infection and trachoma have identified host, chlamydial and environmental factors that are important in heterogeneous responses and outcomes observed in trachoma-endemic communities. Partial immunity does develop and is demonstrated by reduced frequency and duration of Ct infections and inflammatory disease with age. No study has identified a correlate that consistently indicates immunity to Ct infection. To mimic immunity that occurs naturally and represent the variation in the longitudinal data collected, individuals were split based on the frequency and duration of Ct infections during the six-month follow up. Individuals who had no or infrequent infections or were able to resolve within two weeks were considered immune, those who had frequent or long duration infections were considered non-immune. A previous study in The Gambia found that the median duration of ocular Ct infection drops from 3.8 weeks in children aged 0-4 years to 2.1 weeks in those aged

5-14 years⁸. This supports the classification of individuals who resolved infection within two weeks as being partially immune.

This study was designed and undertaken to best control for confounding risk factors for differences in immunity, primarily age, gender and variable history of exposure. The median ages of the protected and susceptible groups were 9 years (2-12) and 8 years (1-12) respectively. Females accounted for 40 % and 43 % of the two groups. At baseline in the study area the ocular Ct prevalence was 20.9 % and the active trachoma prevalence was 21.5 %, indicating meso-endemicity and ongoing transmission¹²⁸. Village membership was different between protected and susceptible individuals, however study sites were selected based on active trachoma prevalence in excess of 20 %. Therefore there was no systematic difference in history of exposure. Results were adjusted for age, gender and village of individuals to eliminate bias. The common demographics of the protected and susceptible groups give confidence that observed differences in antibody responses are due to variable development of immunity.

4.3.2. More focussed antibody responses protect from acquisition and long duration of infection

The breadth and diversity of antibody responses was significantly higher in individuals susceptible to infection. A more focussed response targeted against a smaller number of antigens protected from infection. Lower breadth means generating antibody responses against fewer antigens was protective. Lower diversity means within targeted antigens, making stronger responses against a reduced subset was protective.

Ct is known to stimulate strong B-cell and antibody responses, both systemically and in local sites of infection^{287, 288}. Low-dose Ct infections *in vitro* induce a predominantly B-cell driven response²⁸⁹, ocular infections are lower load than urogenital infections therefore may promote a similar response. The development of neutralising antibodies has been demonstrated *in vitro* and in mouse and non-human primate models^{132, 290, 291}. It is logical that an antibody response targeted against a small number of antigens capable of neutralising Ct infectivity would be protective.

B-cell proliferation and subsequent production of antibodies by plasma cells is resource-limited and also dependent on the levels of the recognised antigen and T-cell help. As B-cells recognise more antigens the availability of these will be reduced,

limiting the generation of new antibody responses. This is equally true for maintenance of plasma cells after initial activation during primary infection. Plasma cells have to compete for survival niches, the ability of these cells to be long-lasting is dependent on the initial levels of antigen-specific plasma cells and how often they encounter and are reactivated by the relevant antigen²⁹². In individuals susceptible to Ct infection the development of antibody responses against a broad range of antigens may limit the initial proliferation and subsequent persistence of potentially protective antigen-specific plasma cells while simultaneously wasting resources on targeting of irrelevant and non-protective antigens.

Partial immunity to trachoma increases with age, therefore it requires time and presumably repeated exposures to Ct to progress. Development of neutralising antibodies as part of this immunity would require repeated reactivation of antigen-specific plasma cells and affinity maturation of the targeting antibodies, as seen with neutralising antibodies in HIV²⁹³⁻²⁹⁵ and non-human primate models of trachoma¹³⁶. Affinity maturation is the process by which antibody affinity to its specific inducing antigen increases. This involves somatic hypermutation of variable regions within surface-bound antibodies on B-cells, which are then selected for based on their new affinity for the inducing antigen which are trapped and presented within lymphoid follicles. In susceptible individuals where antibody responses have been produced against a greater number of antigens it follows that high levels of diverse antigens will be presented within follicles. This means B-cells with lower affinity will have a greater chance of binding antigen and therefore will survive and proliferate. This would weaken the power of selection to remove weak-binding antibodies and would maintain the initial diverse antibody profile. Enhanced survival and expansion of a large pool of antigen-specific B-cells would also put a strain on the resources described above.

4.3.3. Higher responses to 42 antigens were associated with susceptibility to infection

Since diverse antibody responses were shown to increase susceptibility to infection, it was expected that protected individuals with partial immunity would produce stronger responses against a small number of potentially protective antigens. Univariate analyses identified 42 antigens which were differentially recognised between protected and susceptible individuals, however higher responses against these antigens were

associated with susceptibility. This means within the previously discussed diverse antibody profile, non-immune individuals also produce strong responses which individually associate with a lack of protection. The majority of these antigens remained significantly associated with susceptibility after adjusting for current infection or active disease, therefore they were not markers of ongoing infection or inflammatory environment. Antibody responses against these antigens were increasing susceptibility to infection.

In vitro studies have shown that antibodies against specific Ct antigens can block the activity of other Ct antigen-specific antibodies. Crane *et al* showed that PmpD antiserum could neutralise Ct infectivity in ocular, urogenital and LGV serovars¹³⁴. Pre-incubation of Ct EBs with antibodies specific to the immunodominant antigens MOMP and LPS effectively blocked the neutralising ability of PmpD, addition of these antigens following PmpD incubation did not impact neutralisation. This shows that the order in which the host produces antibodies and how they are able to contact Ct impacts their protective capacity. A similar phenomenon has been described in the fungal pathogen *Candida albicans*, antibodies against immunodominant cell-surface antigens blocked the protective effect of antibodies against underlying β -glucans in the inner cell wall²⁹⁶. Three of the 42 antigens associated with susceptibility are localised to the Ct outer membrane; CT017 (Ctd1), CT541 (MIP) and CT579. MIP and CT579 are immunodominant antigens^{297, 298}, high levels of antibodies against these antigens in susceptible individuals may bind to the surface of Ct. This could block the binding and neutralisation of infectivity induced by antibodies against PmpD or other targets of protective antibodies. MIP antibodies can also neutralise infection *in vitro*²⁹⁷, implying it may function similarly to MOMP which has some serovar-specific neutralising epitopes but can block protective antibodies as described above.

Ctd1 has not been previously identified as immunodominant, it could still block binding of protective antibodies. Ctd1 is known to be involved in EB attachment and induction of host-cell signals required for invasion²⁹⁹. It is plausible antibodies binding Ctd1 could improve CT infectivity and promote longer survival in individuals, increasing susceptibility to frequent and longer duration infections.

The majority of the 42 antigens preferentially recognised in non-immune individuals are not localised to the outer membrane, therefore they must have a different method of increasing susceptibility to infection. A large pool of Ct antigens appears to be accessible by the host immune system. Antibodies targeting the majority of these are

not protective. Heightened responses to a diverse panel of these irrelevant antigens divert the humoral immune response, or at least a significant portion of it, away from protective antigens and epitopes. The frequent recognition of these non-protective antigens, demonstrated by their strong associations with susceptibility, means presentation of these antigens to the host may be a deliberate evasive tactic by Ct. Ct readily presents these antigens to the host immune system to serve as decoys. Individuals who do not frequently recognise these irrelevant antigens are better protected from Ct infection. In this 'decoy' hypothesis individual antigens would not be expected to be capable of predicting susceptibility to infection, because this evasion tactic relies on a broad antibody response to limit development of protective antibodies. In susceptible individuals, a broad antibody response with heightened responses to a number of Ct antigens associated with a lack of protection.

4.3.4. Presentation of non-surface antigens to the humoral immune system

An unanswered question from the described 'decoy' hypothesis is how Ct is able to present a large number of non-surface antigens to the immune system. Ct induction of non-surface antigens is not a novel finding, protein-based arrays of sera from patients with urogenital Ct infection have had similar results and *in vitro* studies from the 1980s often described a high prevalence of antibodies that do not bind the EB surface. Bard and Levitt determined that less than 1 % of the total antibody response induced by Ct was targeted against EB surface antigens²⁸⁷. The *in vitro* defined Ct developmental cycle provides some clues as to the source of non-surface antibodies³⁰⁰.

Infectious EBs are adapted to resist environmental stresses but they cannot survive outside cells for a long time. Approximately 40 % of EBs are viable 4 hours after host cell lysis and only 3 % by 24 hours¹⁰¹. Survival time is improved by extrusion of EBs to 76 % at 4 hours and 32 % at 24 hours, however *in vivo* the extracellular stage of the Ct developmental cycle is likely to be short. The opportunity for antibodies to bind EB surface antigens is therefore limited, particularly those released within extrusions. In contrast when infected cells lyse any Ct antigen localised outside of the bacteria could be targeted, particularly those secreted into the host cytosol which were over-represented in the 42 susceptibility associated antigens.

Extrusions are the double-membrane structures EBs reside within after extrusion. The membrane of these extrusions seem to be made up of both host plasma membrane and inclusion membrane factors. These extrusions are reported to improve survival and transmission of Ct through phagocytic uptake into macrophages for dissemination, although this effect is more pronounced in LGV strains compared to urogenital and ocular strains⁷². A recent study of extrusions suggested that the outer membrane is entirely derived from the host plasma membrane, with the inclusion membrane inside. The integrity of this membrane in the extracellular space is not known. The electron microscopy images from this study show limited evidence of the double membrane structure¹⁰¹, additionally phagocytosis relies upon recognition of non-self. This suggests inclusion membrane proteins (Incs) may form a part of the surface of extrusions or become exposed prior to infection of new cells. Localisation predictions of Incs are variable, but 5 which have been demonstrated to localise to the inclusion membrane were found to be susceptibility associated. Five out of 42 is a two-fold higher proportion than Incs in the complete proteome, predicted to be 58 out of 890 to 950 open reading frames depending on Ct serovar³⁰¹. Incs incorporated into the membrane of extrusions would be exposed to the extracellular environment for considerably longer than EB surface antigens providing access to B-cells and antibodies.

Ct can also generate membrane vesicles, these are derived from the bacterial cell envelope and are released from within inclusions³⁰². They have been seen in the host cytosol adjacent to the inclusion and extracellularly. They are involved in responses to stress and stimulating innate immunity³⁰³. The composition of these membrane vesicles is not uniform but independent studies have identified Ct immunodominant antigens including MOMP, Pgp3 and CPAF in these structures³⁰⁴. Incs have also been associated with these membrane vesicles, specifically IncA (CT119), IncF (117) and IncG (CT118)³⁰². IncA and IncG were both identified as susceptibility associated, extracellular release through membrane vesicles may provide another mechanism for antibody responses to target these and other antigens.

4.3.5. Limited identification of antigens associated with protection

This study identified antibody profiles and individual responses which enhance susceptibility to Ct infection and showed that more focussed responses were associated

with protection. There was limited evidence of individual, focussed antibody responses which would support the observed partial immunity. A common feature of human pathogens is the slow development of partial or complete immunity. Repeated exposure to the pathogen is generally the driving force, for example immunity to malaria develops more slowly in areas of lower-endemicity³⁰⁵. This study included children from a trachoma-endemic community aged 1 to 12. While immunity to Ct infection begins to develop from the first exposure, it is possible that development of neutralising antibodies such that they can be detected serologically requires further exposure and subsequent boosting of these antibodies.

The 'decoy' hypothesis proposed may also be involved in limited detection of antigens associated with protection. The results from this study do not support a dichotomy in which individuals with broad antibody responses are completely susceptible and those with focussed responses are entirely protected. This study supports a spectrum of responses where the breadth and diversity of an individual's antibody response is indicative of differential immunity. Protected individuals produce antibody responses against some of the antigens associated with susceptibility, but not to the same extent as non-immune individuals. With repeated exposure protected individuals with more focussed responses are likely to boost protective antibodies while those against the decoys are selected out. In children where immunity is still developing, protective responses are likely produced but diluted down by non-protective responses. In support of this, responses from individuals with a more focussed antibody profile, classified as 'globally low', highlighted 5 antigens in which heightened responses were associated with protection.

How closely *in vitro* protein expression on this array represents native structure is an unknown factor which could have impacted identification of protective responses. Focussing on MOMP and PmpD, they are large, multimeric, membrane-spanning proteins capable of inducing neutralising antibodies *in vitro* and in animal models^{306, 307}. PmpD is also cleaved into a smaller soluble form. The antigenic and neutralising epitopes on these proteins have been well characterised, it is unclear when expressed as monomers on a micro-array whether these epitopes are exposed. PmpD was included in the 441 filtered antigens, MOMP was not and therefore was recognised infrequently. MOMP is an immunodominant protein in urogenital and ocular Ct infection, its role in protection is not fully understood but it should be commonly recognised in this population. It is possible that neutralising antibodies targeting epitopes within MOMP

and PmpD arise infrequently in natural infections. That would not preclude antibodies targeting these antigens from being protective, it would help explain why immunity is slow to arise and only ever partial.

Protein expression and antibody responses from the array were validated by retesting 4 antigens identified, as susceptibility associated using an in-house ELISA. A subset of sera had systematically lower responses to all antigens by ELISA including the positive control Pgp3, these were excluded. CT089 and CT875 were expressed and purified using a histidine-tag. Results from the array and ELISA were strongly correlated. CT118 (IncG) and CT119 (IncA) were synthesised as biotinylated 16mer peptides. Results from the array and ELISA were not correlated. The lack of correlation for IncG and IncA was most likely due to the use of short peptides rather than full-length proteins. All 4 antigens remained associated with susceptibility in the 90 arrayed sera and in the complete set of 130. These results validate this micro-array and the antibody responses identified. Array-based methods are valuable tools for examining global antibody profiles and for target discovery.

4.4. Conclusions and future work

This study identified focussed antibody responses as a correlate of immunity in children who were protected from acquisition of Ct infection or quickly resolved infection. Children who were susceptible to frequent and long duration infection had less focussed antibody responses. Strong antibody responses against 42 Ct antigens were associated with this lack of protection. Ctd1, MIP and CT579 are EB outer membrane proteins that may increase susceptibility to infection by blocking of protective responses or by enhancing Ct attachment and infectivity. The majority of susceptibility associated antigens are non-surface antigens. These are hypothesised to be presented by Ct to the host immune system as non-protective decoys, diverting antibody responses away from protective antigens and epitopes.

Future work should aim to validate the decoy hypothesis and individual surface-exposed antigens through further study of longitudinal antibody responses. A previous longitudinal study in The Gambia identified antibodies against MOMP in tears to be associated with an increased incidence of active trachoma²⁸⁰. Prevalence of active trachoma was high at the time of this study, therefore antibody responses indicative of

the decoy type identified would be expected. This study also demonstrated that tear and serum antibodies may not be correlated in trachoma. Tear samples were collected at every visit for the study described in this section. Micro-array screening should be used for identification of decoy antibody responses through determination of diversity metrics. Use of protein fragments or peptides should be considered for large, multimeric Ct proteins to try to improve reliability of the array. If this is not possible, immunoblots or single-antigen ELISA should be used for antigens likely to be poorly expressed on micro-arrays. Micro-array screening with tear samples would be difficult because antibody levels are generally low and therefore hard to detect. Individual targets which were associated with susceptibility should be tested using tear samples by ELISA to see if they correlate with serum antibody responses and to determine the impact of incident Ct infection on these antibodies. Finger-prick blood samples were collected at the end of a more recent 4-year longitudinal study of children in Tanzania (chapter 3.4.6). Individual antigens should be tested in these samples to investigate if the antibody-based correlates of immunity identified here are present in a geographically and culturally distinct trachoma-endemic population.

5. Global profiling of *Chlamydia trachomatis*-specific antibody responses in Trachomatous Trichiasis

5.1. Introduction

5.1.1. Progressive scarring in trachoma and associated risk factors

The previous chapter examined the antibody profile of children in a trachoma-endemic community to try and identify correlates of humoral immunity which could limit and resolve ocular Ct infection. Clinical signs of inflammation as a result of Ct infection are significant risk factors for progression from active trachoma to conjunctival scarring^{273, 308-310}. Deposition of scar tissue in the conjunctiva can cause deformation of the eyelid, causing it to tighten and be pulled inwards causing the rim of the eyelid and/or the eyelashes to contact the surface of the eye (entropion/trichiasis). Trichiasis can cause pain and physical abrasions to the cornea with the potential for corneal opacities and blindness¹⁴³.

The majority of people in trachoma-endemic communities do not progress to these latter stages of trachomatous disease and levels of pathology differ significantly within those that do progress. This heterogeneity seems in part due to the impact of prolonged infection and inflammation, however other risk factors have also been identified³¹⁰. Increasing age is associated with progression both to scarring and worsening of scarring pathology^{271, 310, 311}. Females are generally found to be at greater risk of developing scarring, however continued progression of established scarring seems to be independent of gender^{279, 310}. The presence and quantity of other bacteria, both commensal and pathogenic, is also associated with scarring trachoma^{27, 146, 312}. This is believed to be due to and subsequently playing a role in continued inflammation. Damage to the conjunctival tissue, specifically goblet cells³¹³, in scarring can cause a decline in tear secretion leading to increased tear film instability. This is known as dry eye and can itself cause damage to the conjunctival tissue worsening scarring progression^{314, 315}.

5.1.2. Dysregulation of inflammation and wound healing promote scarring

As described in chapter 1.3, chronic inflammation that results in conjunctival scarring is driven by dysregulation of both innate and adaptive immune responses. Scarred

conjunctival tissue is marked by inflammatory cell infiltrate, disruption and accumulation of collagen and numerous cytokines and host factors involved in tissue remodelling^{127, 143, 145, 316, 317}. The differential expression of microRNAs involved in regulation of inflammation and tissue remodelling highlight the importance of these processes in development and progression of scarring in trachoma⁵⁵. These innate factors involved in conjunctival scarring are likely enhanced by the aforementioned non-chlamydial bacteria.

Dysregulation of T-cell responses has also been associated with trachomatous scarring^{318, 319}. This was initially thought to be driven by delayed-type hypersensitivity (DTH) reaction against Ct antigens, most prominently HSP60¹³⁸⁻¹⁴⁰. In opposition to this, Th1-type lymphoproliferative responses appear to be depressed in adults with scarring¹⁴¹. HSP60 stimulation induced IL-4 from T-cells of scarring adults, suggesting weakened Th1 responses may be involved in scarring pathology³¹⁹. Evidence from conjunctival transcriptomics have implicated both regulatory T-cells and Th17 cells in scarring trachoma^{142, 309}, supporting regulation of T-cells as important in pathology.

5.1.3. Association between scarring and antibody responses

IgG antibodies against Ct elementary bodies (EBs) have been shown to be significantly higher in individuals with scarring trachoma^{138, 141, 266}, as frequent and persistent infections are associated with scarring this suggests development of these antibodies does not protect from progression. One of these studies found an identical result for the Ct antigen HSP60 independent of antibody responses against EBs, implying it is not simply a marker of increased exposure. This association has not been found consistently in further studies of scarring trachoma and trachomatous trichiasis (TT)^{139, 140}, however one of these studies did find IgG antibodies against another Ct antigen, CPAF, were significantly increased in TT¹⁴⁰. Antibody responses have also been implicated in scarring pathologies resulting from urogenital Ct infection, similarly against EBs and certain antigens including HSP60^{167, 168, 320, 321}.

It is unclear how common these scarring-associated antibody responses are in trachoma endemic communities and if they are directly involved in the scarring process. It is possible via opsonisation that Ct-antibodies could facilitate greater Ct infectivity in early life promoting frequent and prolonged infections that are known to be a risk factor

for scarring. Equally they may be coincidental serological markers of infection. Serum from individuals with scarring trachoma most of which also had TT, were used to screen a Ct D/UW3 proteome array of 908 genomic and plasmid ORFs to elucidate the role of antibodies in scarring trachoma.

5.1.4. Study design and initial analysis

The study was performed as described in chapter 3.2.1 and Lu *et al*¹⁶². Briefly, 61 cases of scarring trachoma with trichiasis (TT) and 61 age, sex and location-matched controls with normal healthy eyes were collected in The Gambia between May 2006 and February 2009. Thirty-four cases and 25 controls were screened on the serovar D Ct proteome array as described previously^{161, 169}, performed on 96 well microtitre plates. This study identified ten antigens that were recognised by over half of the 59 samples tested, of greater interest was the discovery of four antigens preferentially recognised by those with trichiasis and eight recognised by healthy controls (Table 5.1). These included CT117 (IncF), CT442 and CT556 which are *Chlamydia*-specific as well as being reportedly immunogenic and protective¹⁶². IncF and CT442 are both inclusion membrane proteins. IncF has a putative role in bridging interactions between RBs and the host cytosol^{163, 322} and CT442 had previously been identified as immunogenic in women with urogenital Ct infection¹⁶⁹. However, there were a number of limitations in the study both in terms of the generation of the laboratory data and the statistical analysis of the data.

The recombinant proteins on the micro titre array were not purified to a high standard, expression of the target protein was confirmed but various other proteins and/or fragments were also present. This impurity meant the proteins were not quantified prior to use on the array, thus the amount and specificity of protein bound on the plates was unclear. Capturing the GST-fusion proteins with immobilised glutathione and limited background correction would have reduced some non-specific binding but not all and not in a repeatable way. In addition it is unclear how the large GST-expression tag (approximately 26 kDa) would have impacted native confirmation of the proteins^{164, 165}. Certain chlamydial antigens known to induce antibody responses in ocular Ct infection and trachoma were not commonly recognised in this screen, notably MOMP (CT681), HSP60 (CT110), OmcB (CT443) and most polymorphic membrane

proteins (Pmps). It is possible the purity and protein folding issues described above may have impacted their recognition as they are large, surface exposed proteins *in vivo* with complex tertiary and quaternary structures possibly required for antibody binding^{306, 307, 323, 324}. Except for MOMP these proteins are also variably processed to soluble secreted forms, further complicating understanding of how and in what form the immune system recognises these antigens^{306, 325, 326}.

Aside from these technical issues there were also problems with sample selection and analysis. Out of 61 matched trichiasis cases and healthy controls available only 59 with the highest antibody titre against Ct ocular serovar EBs (A – C) were screened on the complete array, 34 cases and 25 controls. This reduced the number of samples tested and therefore the amount of natural variation in responses was insufficiently examined. By filtering on responses to EBs, essentially a small pool of immunodominant antigens, people with non-classical or differentially targeted immune responses may have been excluded. Common methods for analysing array-type data were also not utilised. No attempt was made to normalise across the 908 antigens, despite this results were analysed using tests that assume normality. No correction for multiplicity of testing was applied and false positives were not accounted for¹⁶⁶. The expected number of false positives was calculated using only the 19 antigens which had statistically significant ($p\text{-value} \leq 0.05$) associations rather than the complete 908, which from their calculations would have meant 45 potential false positives. These issues highlight the need for validation both within this Gambian population and in other trachoma-endemic communities.

A more robust analysis of results from this array including examination of global patterns of antibody responses, similar to Chapter 4 was performed. The array as a whole and specific immune targets associated with scarring or healthy controls were subjected to further independent serological testing.

Table 5.1. Patient demographics in protected and susceptible groups.

Associations were determined using a generalised linear model.

	HEALTHY CONTROLS	SCARRED CASES	ASSOCIATED P-VALUE
NUMBER	25	34	NA
AGE IN YEARS (95% CI)	60.00 (32.40-75.00)	61.00 (34.83-80.87)	0.357
FEMALE (N [%])	19 (76.00)	21 (61.76)	0.251

5.2. Results

5.2.1. Array Normalisation and filtering

The raw optical density intensity data was slightly positively skewed, p-value 0.018 (Figure 5.1A). Transformation and normalisation of the array was assessed using relative rank deviation (RRD)²⁰⁶. The most suitable data transformation technique was inverse hyperbolic sine transformation, which reduced RRD primarily in antigens with the highest signal intensities (Figure 5.1B). Mean-centring and trimmed means of 2 and 10% were tested to normalise the data. A trimmed mean of 2% performed better than 10%, and mean-centring reduced the RRD compared to both (Figure 5.1C).

Antigens were filtered by excluding any with mean optical density lower than the global mean across all antigens, this removed infrequently recognised antigens which would add noise to the data. The normalised data of 59 serum samples were then analysed against 230 antigens. This data set was still positively skewed, p-value 0.020, so non-parametric tests were deemed more appropriate and used throughout (Figure 5.1D).

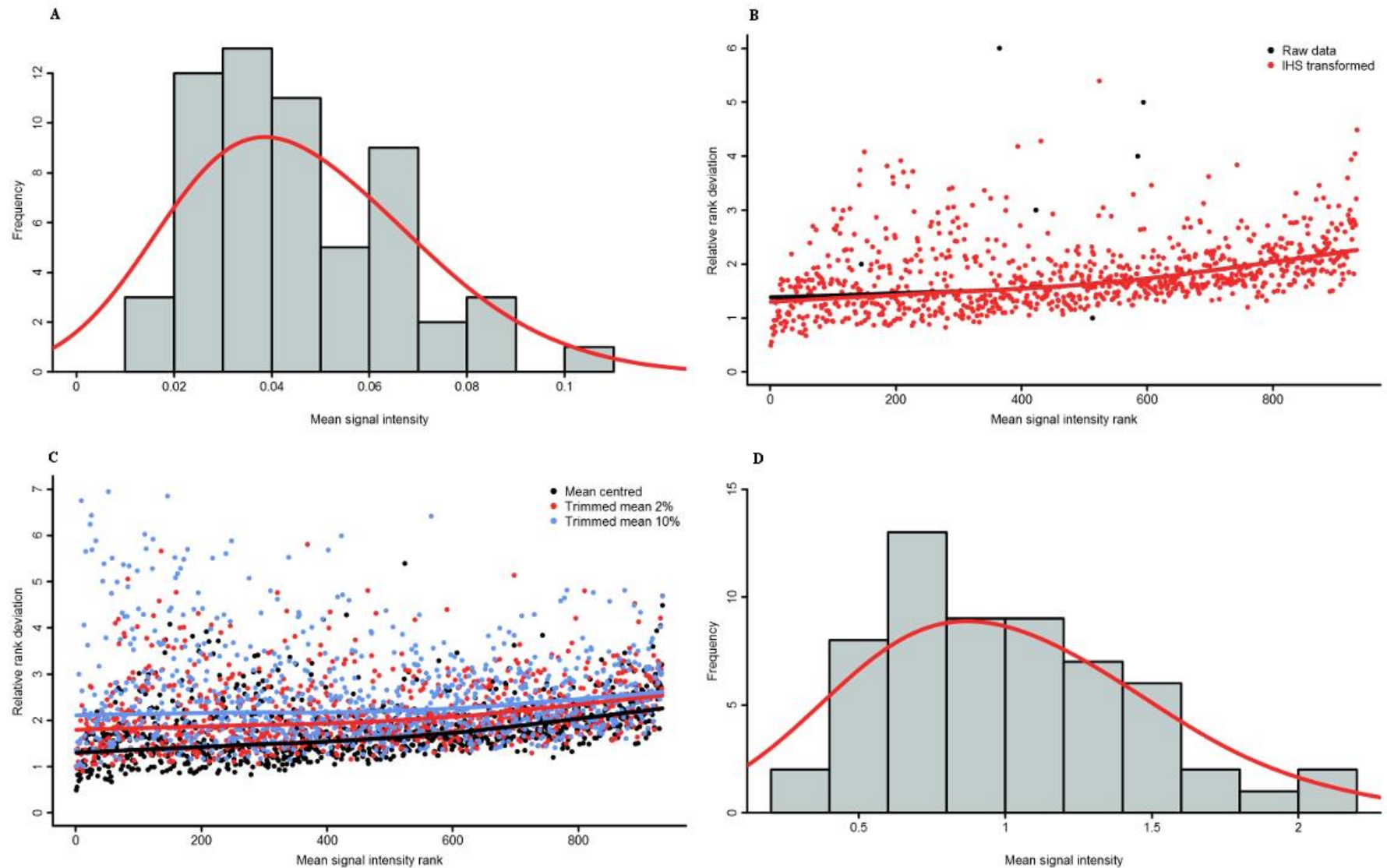


Figure 5.1. Transformation and normalisation of the raw array data.

A) Histogram showing positively skewed distribution of the raw data. B) Significantly reduced deviation (RRD) in the data after inverse hyperbolic-sine transformation (red) compared with the raw data (black). C) Reduced deviation in the data after mean-centring (black) compared with normalisation by trimmed-mean at 2 % (red) and 10 % (blue). D). Histogram of normalised data still showing a slight positive skew.

The normalised continuous data was used for all possible analyses, for methods which required categorical data samples were divided into groups on a per antigen basis as described in the previous study. Binary classification into positive and negative samples per antigen was the desired outcome and empirically was better than all other cluster sizes up to ten, $p\text{-value} < 0.006$ (Appendix Figure 3). K-medoids clustering was the worst performer for a cluster size of two across all 230 antigens. The remaining methods performed well, a median average silhouette width between 0.7 and 0.8, with hierarchical clustering performing best on average (Figure 5.2). Clustering method for each antigen was selected using their highest individual average silhouette width from the five methods, this was significantly better than using one method for all, $p\text{-value} < 0.011$, except for hierarchical clustering, $p\text{-value} 0.385$ (Figure 5.2).

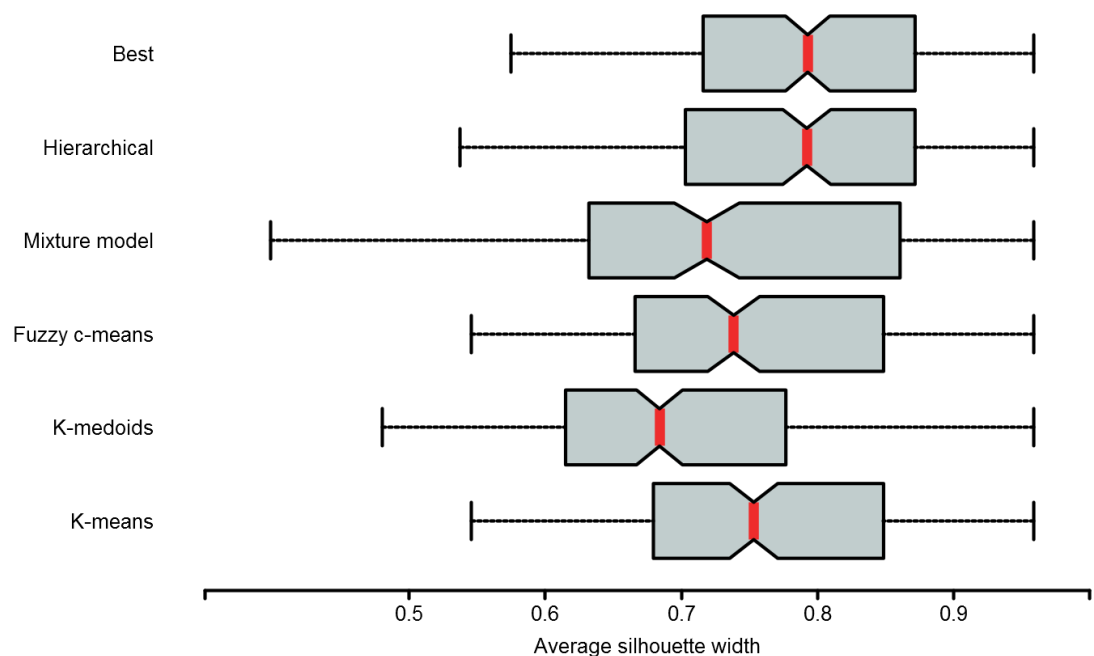


Figure 5.2. Average silhouette widths for clustering method trialled for all 441 antigens.

'Best' method had the highest median across all antigens. Clustering methods are detailed on the left-hand side. Red lines indicate the median. Notches were calculated as $\text{median} \pm 1.57 \times \text{IQR} / \sqrt{n}$, where IQR is the interquartile range and n is the number of samples. The whiskers were calculated by adding 1.5 times the IQR to the 75 percentile and subtracting 1.5 times the IQR from the 25 percentile.

5.2.2. Antibody responses were focussed on proteins expressed early and late during the developmental cycle and localised to interact with the host

Comparing the 230 antigens remaining after filtering with the 908 proteins screened on the array, there was a significant over-representation of genes whose peak expression is either very early or very late in the developmental cycle, likely representing antigens important in entry and those exposed at exit, p-values 0.003 and 0.025 (Figure 5.3). Proteins predicted to be extracellular/secreted, or localised to the outer membrane and periplasm were also over-represented in the immunogenic antigens, p-value < 0.001 (Figure 5.4).

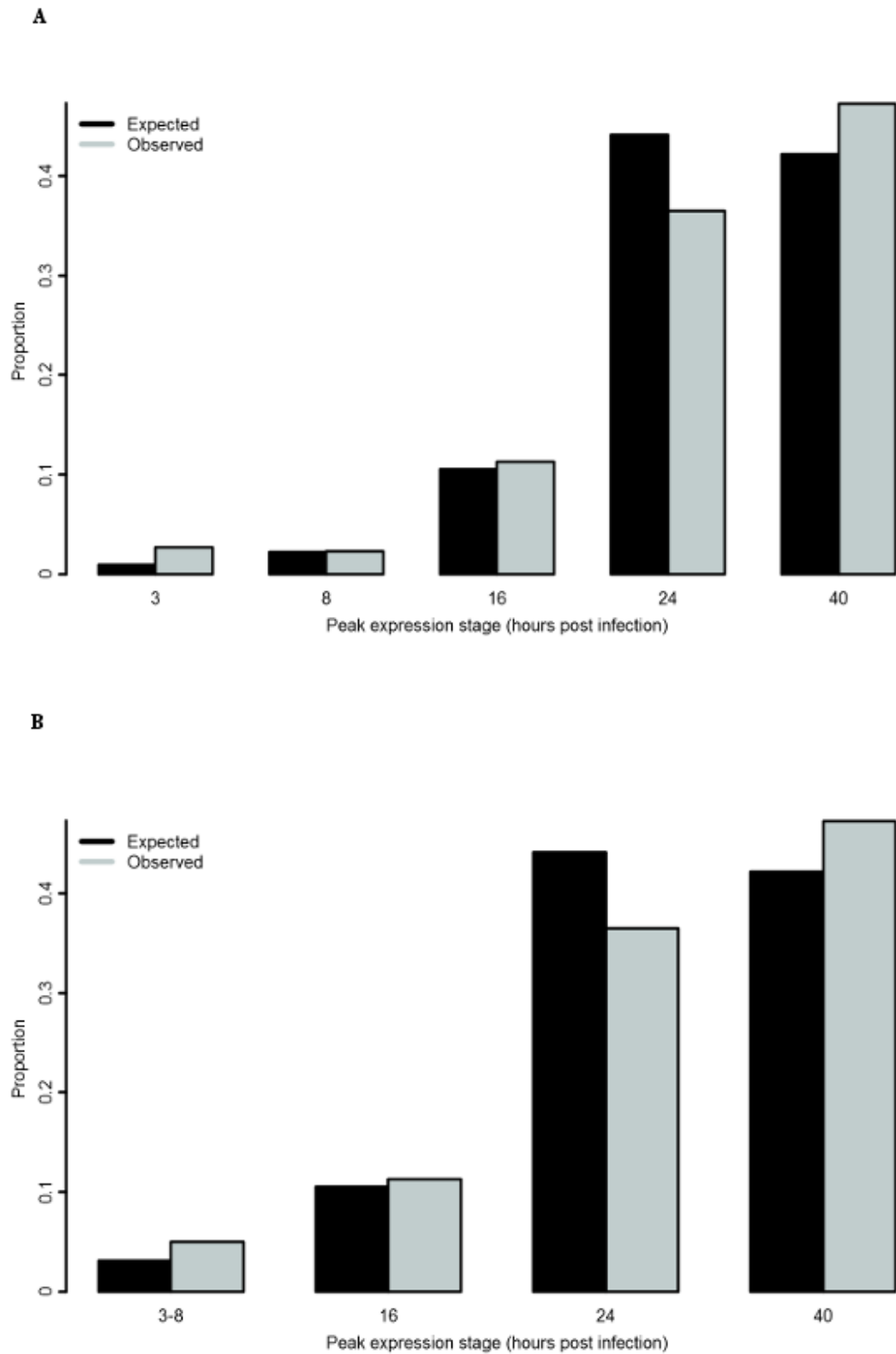


Figure 5.3. Over-representation of late and very early expressed proteins in immunogenic antigens.

Proteins identified through transcriptomics as expressed late or very early in the Ct developmental cycle were over-represented in the 230 immunogenic antigens (grey) compared with the total unfiltered 908 (black).

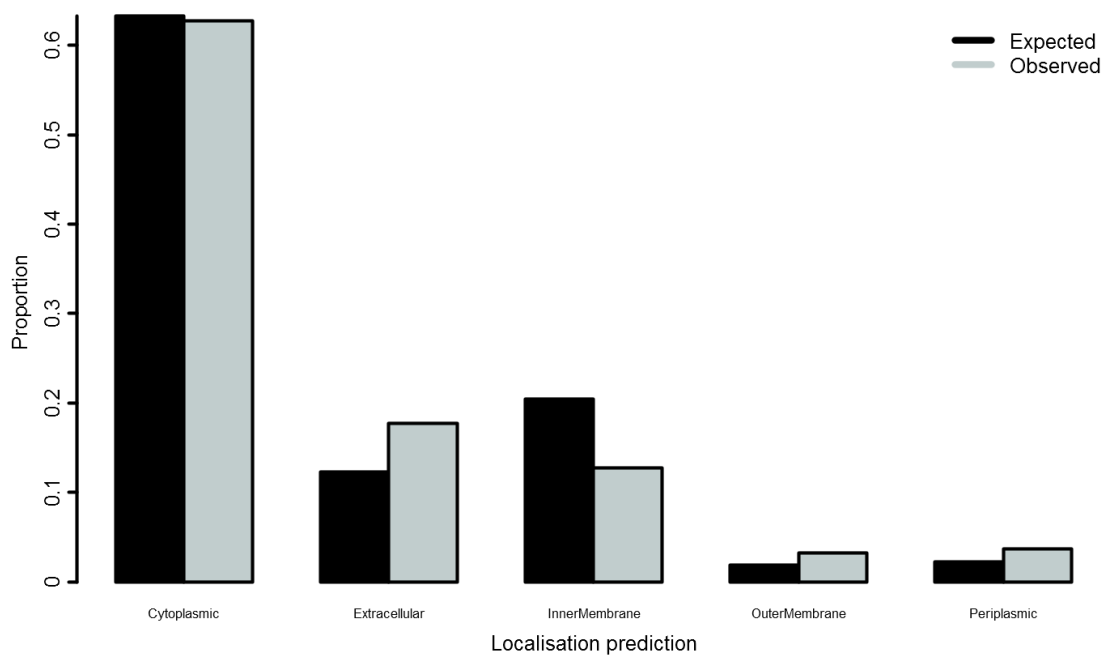


Figure 5.4. Over-representation of proteins extracellular, outer membrane and periplasm in immunogenic antigens.

Proteins with a consensus localisation prediction of extracellular, outer membrane and periplasm were over-represented in the 230 immunogenic antigens (grey) compared with the total unfiltered 908 (black).

Next the global profile of antibody responses was determined in scarred and healthy individuals. Breadth of response tended to be higher in scarred individuals however it was highly variable and did not reach significance, p-value 0.620 (Figure 5.5).

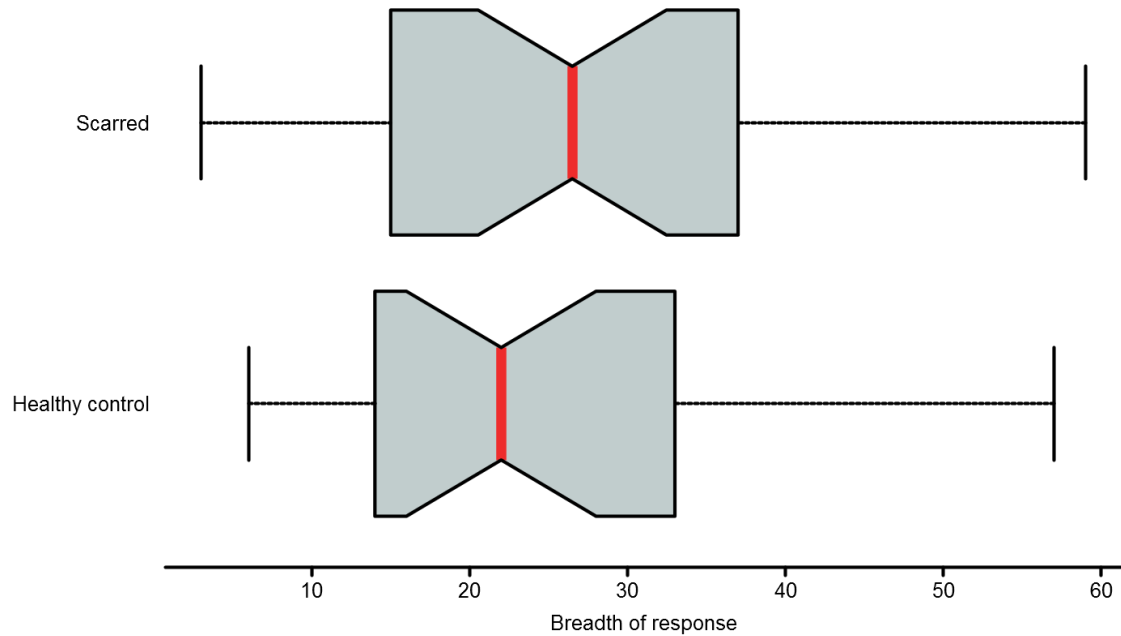


Figure 5.5. No difference in breadth of antigens recognised between adults with and without scarring.

Measured as the number of antigens an individual had a positive response to.

Diversity measured as Hill numbers dropped significantly with increasing order of diversity supporting uneven antibody profiles with a few immunodominant targets (Table 5.2), there was no difference between scarred and healthy individuals. Simpson's diversity index was skewed by a few individuals with very focussed responses and showed no difference between groups, p-value 0.451 (Figure 5.6A). Shannon's diversity index, similar to measures of breadth, tended to be lower in healthy individuals but the difference was not significant, p-value 0.130 (Figure 5.6B).

Table 5.2. Uneven antibody responses in adults with and without scarring, as determined by Hill numbers.

Associations were determined using a generalised linear model.

ORDER OF DIVERSITY	HEALTHY MEDIAN (IQR)	SCARRED	P-VALUE
2	47044.51 (19236.44-1258563.84)	43671.46 (12028.61-197223.87)	0.943
3	3195.34 (1615.14-37337.26)	3020.98 (1148.38-9312.23)	0.719
4	1303.18 (708.39-11569.11)	1240.12 (524.87-3368.23)	0.639
5	832.37 (469.29-6441.03)	794.55 (354.85-2025.98)	0.599

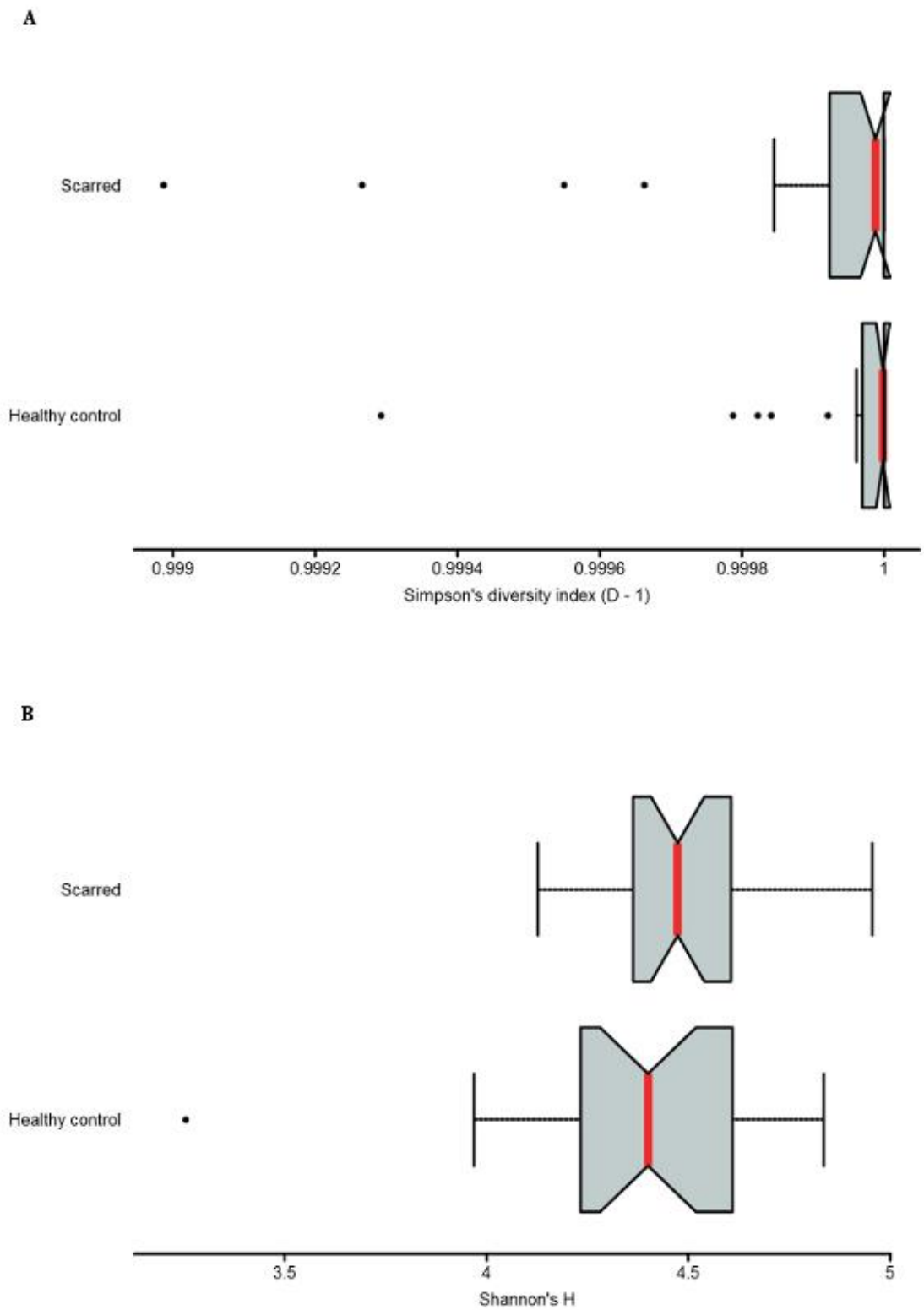


Figure 5.6. No significant differences in diversity of antibody responses between adults with and without scarring.

Diversity was measured using A) Simpson's diversity index and B) Shannon's diversity index.

5.2.3. Individual antibody responses are associated with cross-sectional evidence of conjunctival scarring

Association of responses to individual antigens with evidence of conjunctival scarring was determined using a generalised linear model adjusting for age and gender of the individuals. Nine differential antibody responses were identified between adults with and without scarring, p-values ≤ 0.1 were included to include antigens outside the 95 % distribution that were close to the 0.05 threshold for significance (Table 5.3). All antigens except CT442 had higher responses in scarred individuals (Figure 5.7 parts A and B). None of the differentially recognised antigens had good predictive value.

Table 5.3. Differentially recognised antigens between adults with and without scarring.

Univariate associations were determined using a generalised linear model. Variables were resampled 10,000 times and remodelled to determine permuted p-values (P*).

ANTIGEN	P-VALUE	P*	T	SE (T)	OR	95% CI	AUC
CT667	0.013	0.005	0.78	0.32	2.19	1.28-4.43	0.64
CT645	0.036	0.029	0.56	0.27	1.76	1.09-3.16	0.63
CT314	0.040	0.012	0.42	0.21	1.52	1.09-2.43	0.66
CT698	0.049	0.037	0.55	0.28	1.72	1.05-3.16	0.59
CT471	0.051	0.023	0.59	0.30	1.80	1.08-3.49	0.62
CT442	0.054	0.019	-0.12	0.06	0.89	0.77-0.98	0.63
CT679	0.057	0.011	0.66	0.35	1.94	1.11-4.28	0.61
CT425	0.070	0.049	0.51	0.28	1.66	1.04-3.16	0.54
CT706	0.098	0.064	0.32	0.19	1.37	1.03-2.19	0.57

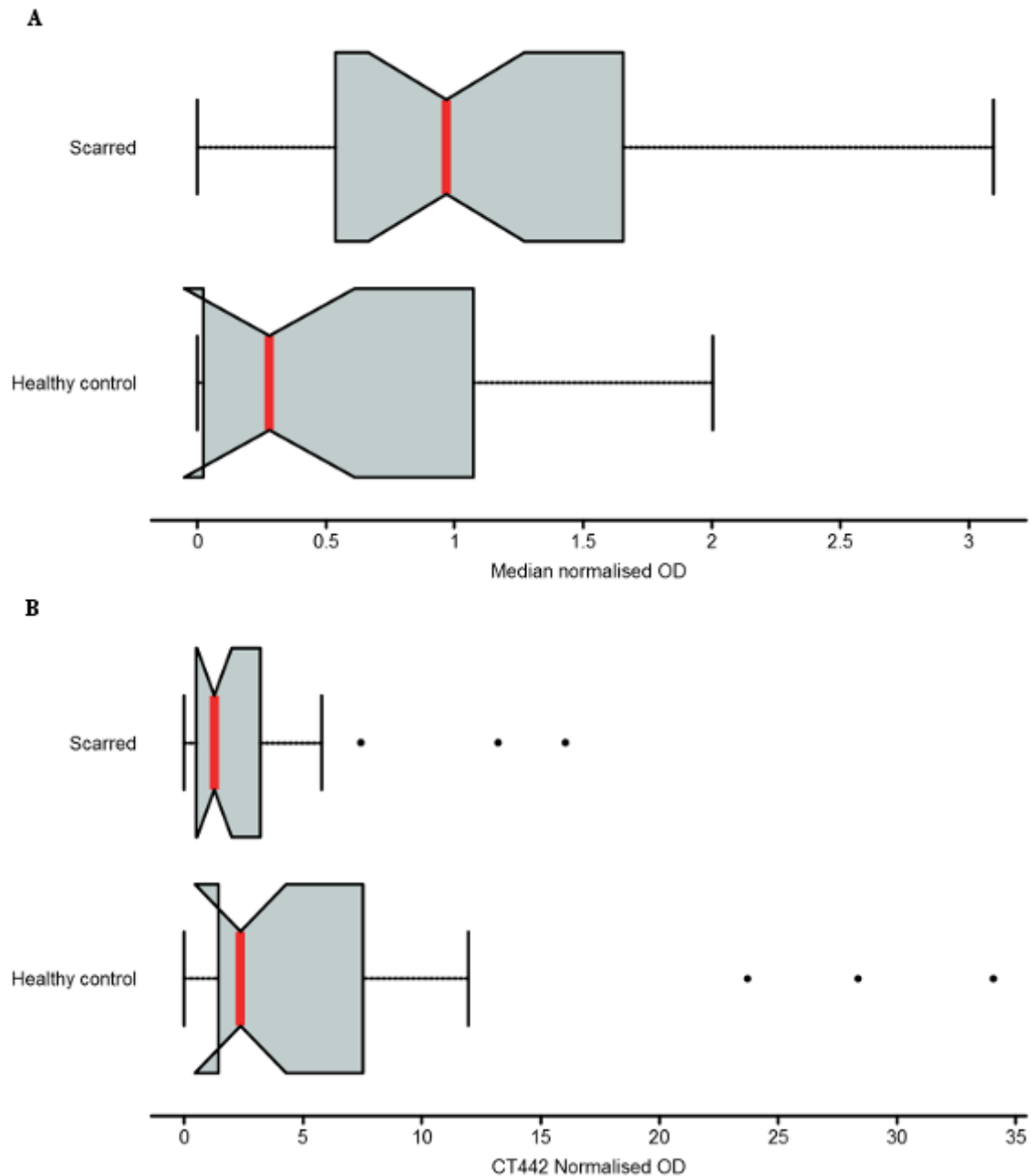


Figure 5.7. Differential antibody responses between adults with and without scarring. Individual responses were averaged across the 8 scarring associated antigens (A) and the single antigen associated with lack of scarring (B).

When combined these 9 antigens were significantly better at predicting clinical outcome, at each step the model was significantly better than a model including just age and gender, p -values < 0.002 (Table 5.4). The model with the strongest predictive value included all nine antigens, however this was not significantly better than use of seven or eight antigens in combination (Figure 5.8A). The predictive value of the multivariate model additively including univariate associated antigens was significantly better than models using randomly selected antigens from one to nine antigens (Figure 5.8B). The

specificity and sensitivity of the model improved with addition of antigens, 9 antigens had 100 % specificity and 47.06 % sensitivity. At 100 % sensitivity these were only 48.00 % specific (Figure 5.8C). A best case scenario balancing specificity against sensitivity from this model yielded 88.00 % specificity and 76.47 % sensitivity. This minimised false positives, which is of greater value in communities where Ct and active trachoma prevalence is declining.

Table 5.4. Combinatorial antibody responses increased predictive power.

Generalised linear models including antigens in a step-wise manner increased as more antigens were included, determined by model quality (AIC) and likelihood ratio test comparing with the null model including only covariates (P-VALUE). Predictive value (AUC) showed modest progressive improvement.

NUMBER OF ANTIGENS INCLUDED	AIC	P-VALUE	AUC
1	76.50	0.0017	0.64
2	76.24	0.0023	0.65
3	75.35	0.0018	0.72
4	76.86	0.0038	0.70
5	77.41	0.0046	0.71
6	72.22	0.0005	0.77
7	71.22	0.0003	0.80
8	71.25	0.0003	0.77
9	71.33	0.0003	0.83

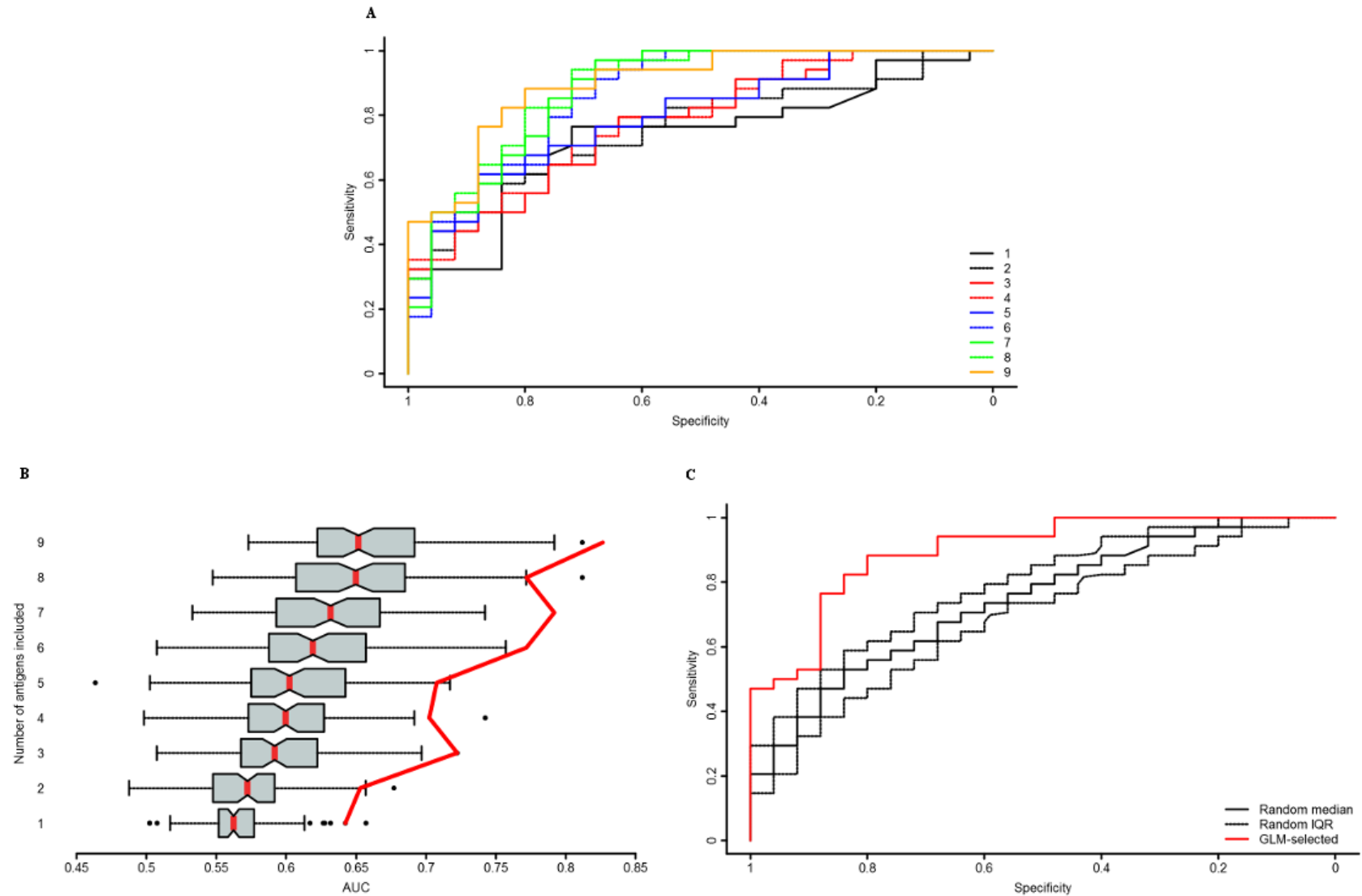


Figure 5.8. Combinatorial antibody responses did not accurately predict presence of scarring.

A) Models including increasing numbers of differentially recognised antigens showed modest increases in specificity and sensitivity. B) Antigens individually associated with scarring were significantly better at predicting scarring than antigens chosen at random, as determined by AUC.

Number of antigens included is indicated on the y-axis. C) The 9 differentially recognised antigens outperformed 9 randomly select antigens when predicting scarring.

5.2.4. Combinatorial antibody responses identified by multivariate regression did not improve predictions of conjunctival scarring

A multivariate random forests regression was performed including all 230 antigens to determine if evaluating all antigens simultaneously identified antibody targets not found in the univariate analysis. Six antigens had variable importance outside 95 % of the normal distribution, measured as the mean decrease in accuracy of trees split on each variable, and a further 3 were within 1 standard deviation of these (Table 5.5 and Figure 5.9). Four of these were also associated with scarring outcome in the univariate analyses. A multivariate model including either 6 or 9 antigens was again significantly better than randomly selected antigens, the model including 9 outperformed the model including 6 based on AUC (Figure 5.10).

Table 5.5. Most important antigens for classification by random forests regression.

A multivariate random forests regression was performed with all 230 antigens. These 9 antigens were in the top 2.5 % based on variable importance.

ANTIGEN	VARIABLE IMPORTANCE
CT471	0.32
CT635	0.25
CT230	0.23
CT695	0.20
CT442	0.19
CT610	0.19
CT314	0.18
CT855	0.18
CT667	0.18

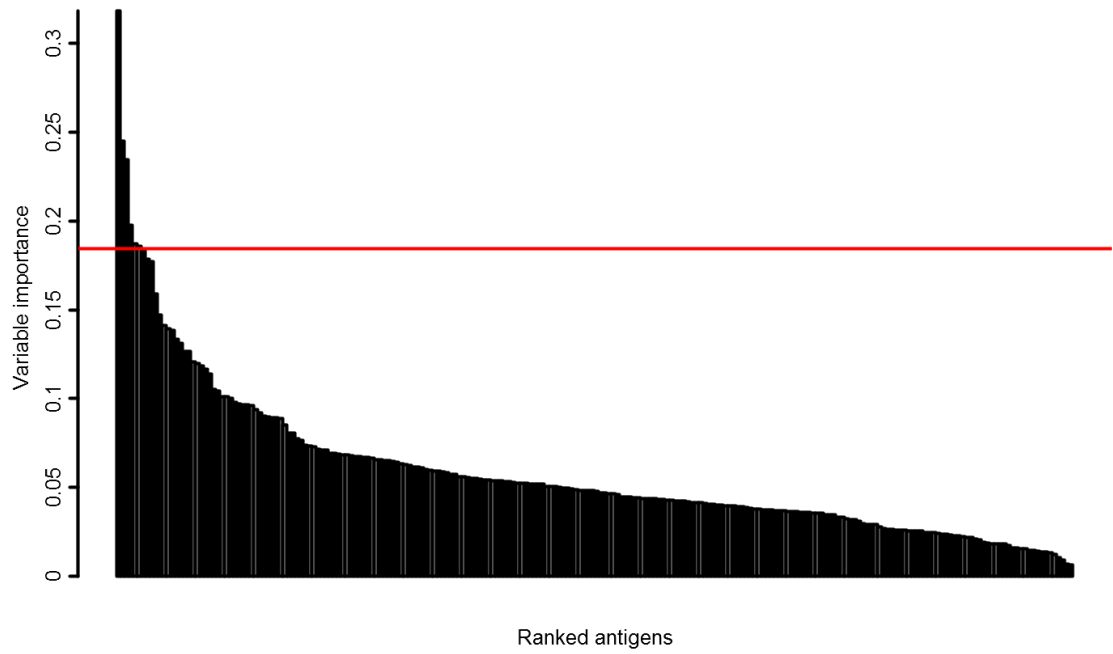


Figure 5.9. Ranked variable importance from random forests regression.

The Gini index was used to calculate variable importance. Antigens in the top 2.5 % of the distribution are indicated by the red line.

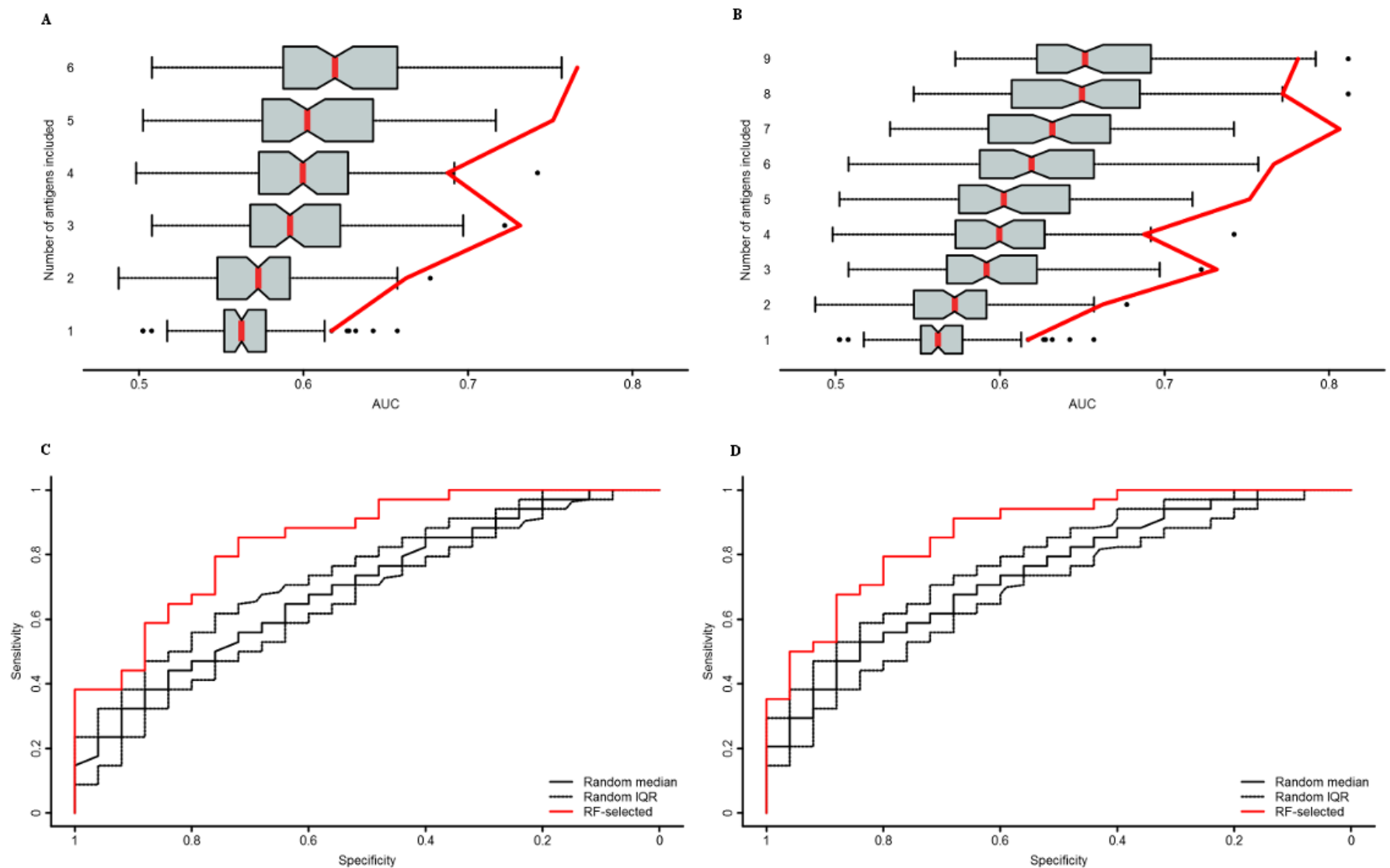


Figure 5.10. Multivariate regression identified antigens with modest predictive power.

The top 6 (A and C) or 9 (B and D) antigens by variable importance from a random forests regression were included in a generalised linear model.

A and B) AUC of the models including 6 or 9 random forests-selected antigens in a step-wise manner were better than a model of randomly selected antigens. C and D) Models of the top 6 or 9 antigens were significantly more sensitive and specific than a model of randomly selected antigens.

5.2.5. Selective expression identified antigens and independent repeat ELISA testing of differentially recognised antigens

To validate the antigens recognised in the microtitre plate array as a whole and individual targets, three antigens were selected for further testing by ELISA. There was no clear pattern of localisation or function in the 9 differentially recognised antigens, for this reason antigens selected for further testing were selected using a combination of homology to proteins of known function and previously published information. CT442 was selected for follow-up because of its potentially interesting intracellular biology as an inclusion membrane protein (Inc) known to induce T-cell responses and the only target with greater responses in healthy individuals¹⁹⁵. CT667 was selected as it is a homologue of CdsG, a conserved bacterial protein involved in type-three secretion (T3S)³²⁷. It's homologue in other bacteria acts as a chaperone for the T3S-needle protein and CT667 has been localised to the host cytosol and around the inclusion membrane, depending on host cell type and stage of the Ct developmental cycle³²⁸. CT706 was selected as a homologue from *C. muridarum* has been previously identified as immunogenic. All 3 were available as GST-fusion constructs. Pgp3 was included as an immunodominant positive control.

CT442 was successfully expressed using the recommended conditions described in chapter 3.2.1, however it was contained in insoluble *E. coli* inclusion bodies (Figure 5.11A). Increasing the time of induction incrementally from 3 hours up to 16 hours and reducing the range of temperature of induction from 30 °C to 16 °C was trialled, but CT442 remained insoluble and expression levels were low (Figure 5.11B). CT442 is known to reside in the inclusion membrane therefore given its predicted structure it is not surprising it was insoluble in aqueous conditions, even when expressed with the GST-fusion which can improve solubility. A panel of zwitterionic and non-ionic detergents were tested to extract the protein from insoluble inclusion bodies. CT442 was partially soluble with ASB-14, DDM and LDAO (Figure 5.11C lanes 4,8 and 10). When this was repeated with a larger culture volume, CT442 was partially solubilised with these detergents. Expression levels were poor and CT442 could not be purified from these samples using glutathione-beads (Figure 5.11D). Co-expression with molecular chaperones as described in chapter 3.1 had a similar result with some increased solubility but low expression levels (Figure 5.11E). It was not possible to purify CT442 when co-expressed with either pKJE7 or pTf16 (Figure 5.11F).

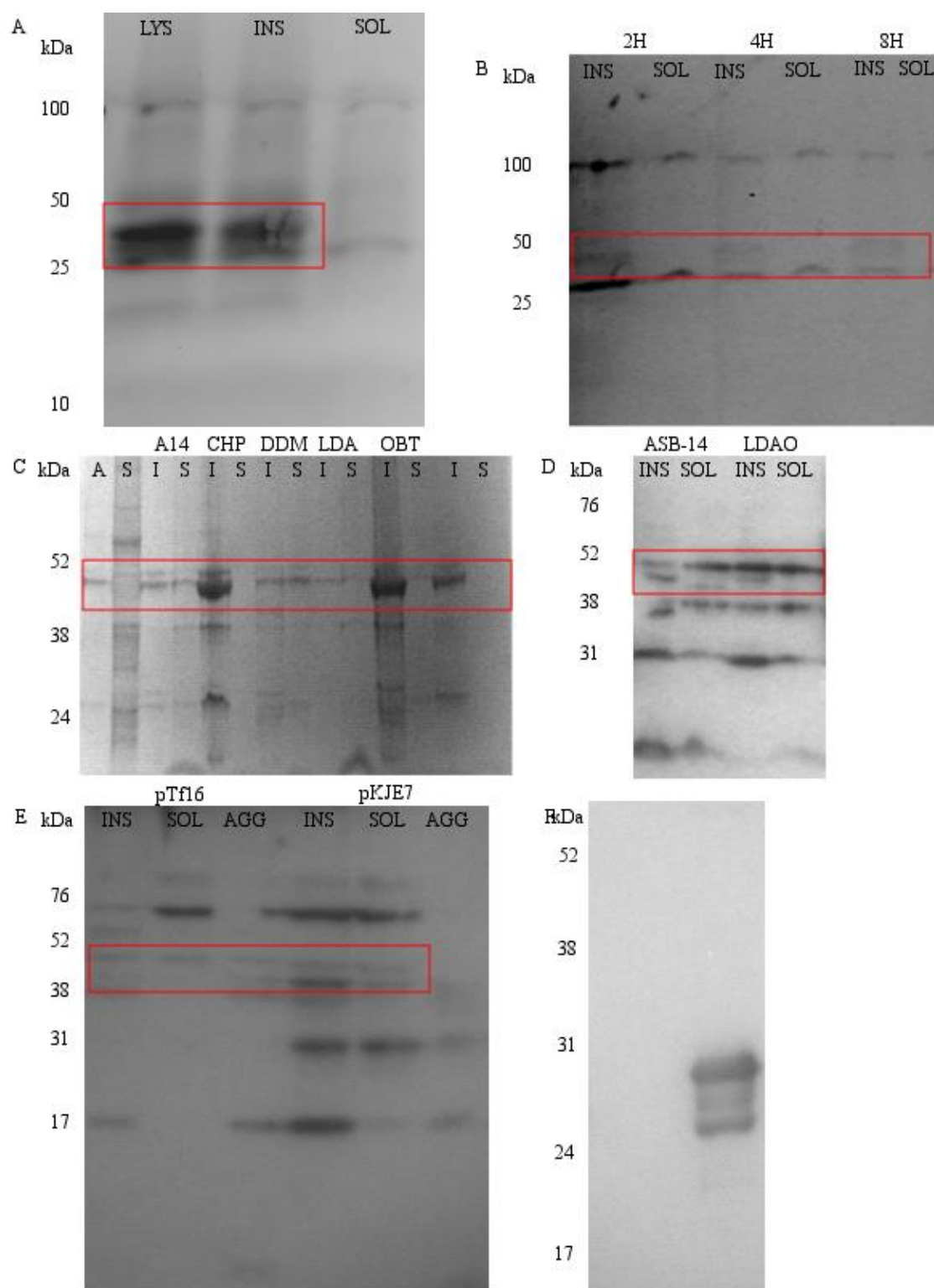


Figure 5.11. Expression trials of CT442-GST.

Western blots were incubated with an anti-GST monoclonal antibody to bind the GST-fusion moiety. The lysates were separated by ultra-centrifugation into aggregate (AGG/A), soluble (SOL/S) and insoluble fractions (INS/I). The relevant bands are highlighted in red. A) Aggregated and insoluble expression of CT442-GST. B) CT442-GST remained insoluble when induced at 16 °C for 2, 4 and 6 hours with 0.2 mM IPTG.

C) CT442-GST was partially soluble in ASB-14 (A14), DDM and LDAO (LDA) but not CHAPS (CHP) or octyl- β -D-1-thioglucopyranoside (OBT). D) CT442-GST expression was low when co-expressed with molecular chaperones, the protein was partially soluble. E) CT442-GST could not be purified from these samples using glutathione beads.

Due to problems with the expression and purification of CT442, expression of two partial-length CT442 constructs, CT442N and CT442C which contained the N-terminal half and C-terminal half of the protein respectively was trialled. Using recommended conditions CT442N expression was undetectable (Figure 5.12A lanes 1-3), CT442C had high expression levels although still primarily aggregated or formed insoluble inclusions (Figure 5.12A lanes 4-6). CT442C did appear to be partly soluble, but the very prominent bands in the neighbouring wells/lanes made it unclear if this was an artefact of the western blot. A detergent screen using only the 3 detergents found to improve solubility of the full-length protein showed ASB-14 was able to partially solubilise CT442C (Figure 5.12B). To reduce aggregation and formation of insoluble inclusion bodies CT442C was induced at 25 °C for 5 hours followed by ASB-14 solubilisation of the insoluble portion. The majority of CT442 was still aggregated, however ASB-14 solubilised nearly all detectable insoluble protein (Figure 5.12C). Purification of soluble CT442C using glutathione-beads from the soluble fraction (Figure 5.12D lanes 1-3) and after ASB-14 solubilisation (Figure 5.12D lanes 4-5) showed mixed success. Most of the protein was unbound, however some was purified successfully from the soluble fraction.

The C-terminal CT442 construct was approximately 16 kDa and the GST-fusion moiety was approximately 29 kDa. The effects of this relatively large tag on recognition of CT442C in serological work was unknown, therefore cleavage of the GST-fusion was performed. CT442C could not be purified after cleavage of the GST-fusion. Due to these problems a CT442 C-terminal peptide, which had been previously shown to be immunogenic (personal communications Bernhard Kaltenboeck) was synthesised as a biotinylated 16-mer peptide. This peptide was produced by thinkpeptides (ProImmune, Oxford, UK) and had a purity of 89.69 %.

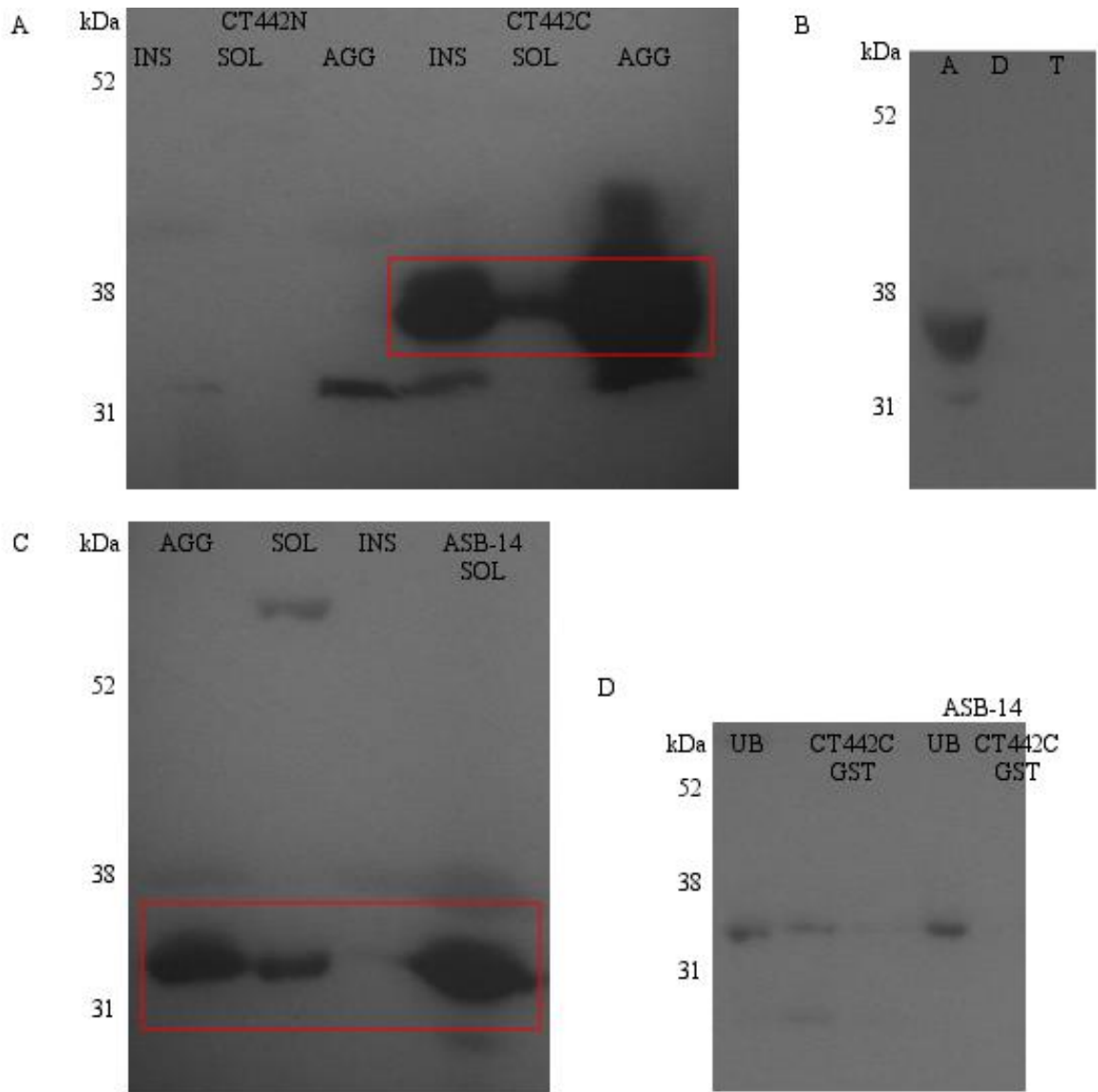


Figure 5.12. Expression trials of CT442N-GST and CT442C-GST.

Western blots were incubated with an anti-GST monoclonal antibody to bind the GST-fusion moiety. The lysates were separated by ultra-centrifugation into aggregate (AGG), soluble (SOL) and insoluble fractions (INS). The relevant bands are highlighted in red. A) Expression of CT442N was undetectable, CT442C was primarily aggregated and insoluble when induced at 30 °C for 3 hours with 0.2 mM IPTG. B) ASB-14 (A) partially solubilised CT442C-GST, DDM (D) and Triton X-100 (T) did not. C) CT442C-GST was partially soluble when induced at 25 °C for 5 hours with 0.2 mM IPTG. D) CT442C-GST was primarily unbound after incubation with glutathione beads (UB), some protein was successfully purified from the soluble fraction (CT442C-GST).

CT667 was soluble using recommended conditions for induction and could be purified with and without cleavage of the GST-moiety (Figure 5.13 parts A and B). Purification after cleavage reduced the number of non-specific bands, however this may have been caused by the reduced concentration of the cleaved CT667 (Figure 5.13C). Size-exclusion chromatography was used to 'clean-up' the purified sample, after which very few non-specific bands were then detectable by silver staining (Figure 5.13D). CT667, following chromatography, was judged to be 'cleaner' by both Coomassie blue and silver staining (Figures 5.13E and 5.13F), however non-specific bands were still present. The most prominent band not corresponding to the expected size of CT667 (16 kDa) was approximately 30 kDa suggesting it may be a dimer rather than a bacterial contaminant.

CT706 was insoluble using recommended conditions rather than aggregated so the ability of a panel of detergents to solubilise the protein was tested (Figure 5.14A lane 1). CT706 was solubilised by a number of detergents. ASB-14 and LDAO were the most effective (Figure 5.14A lanes 2-3 and 8-9). The protein was successfully purified from both of these samples with and without GST-fusion cleavage (Figure 5.14B). More non-specific bands were present in CT706 than in CT667, therefore further purification was needed. Size-exclusion chromatography successfully purified CT706 removing the non-specific bands observed previously (Figure 5.14C). This reduction in impurities was observed by both Coomassie blue and silver staining (Figure 5.14 parts D and E).

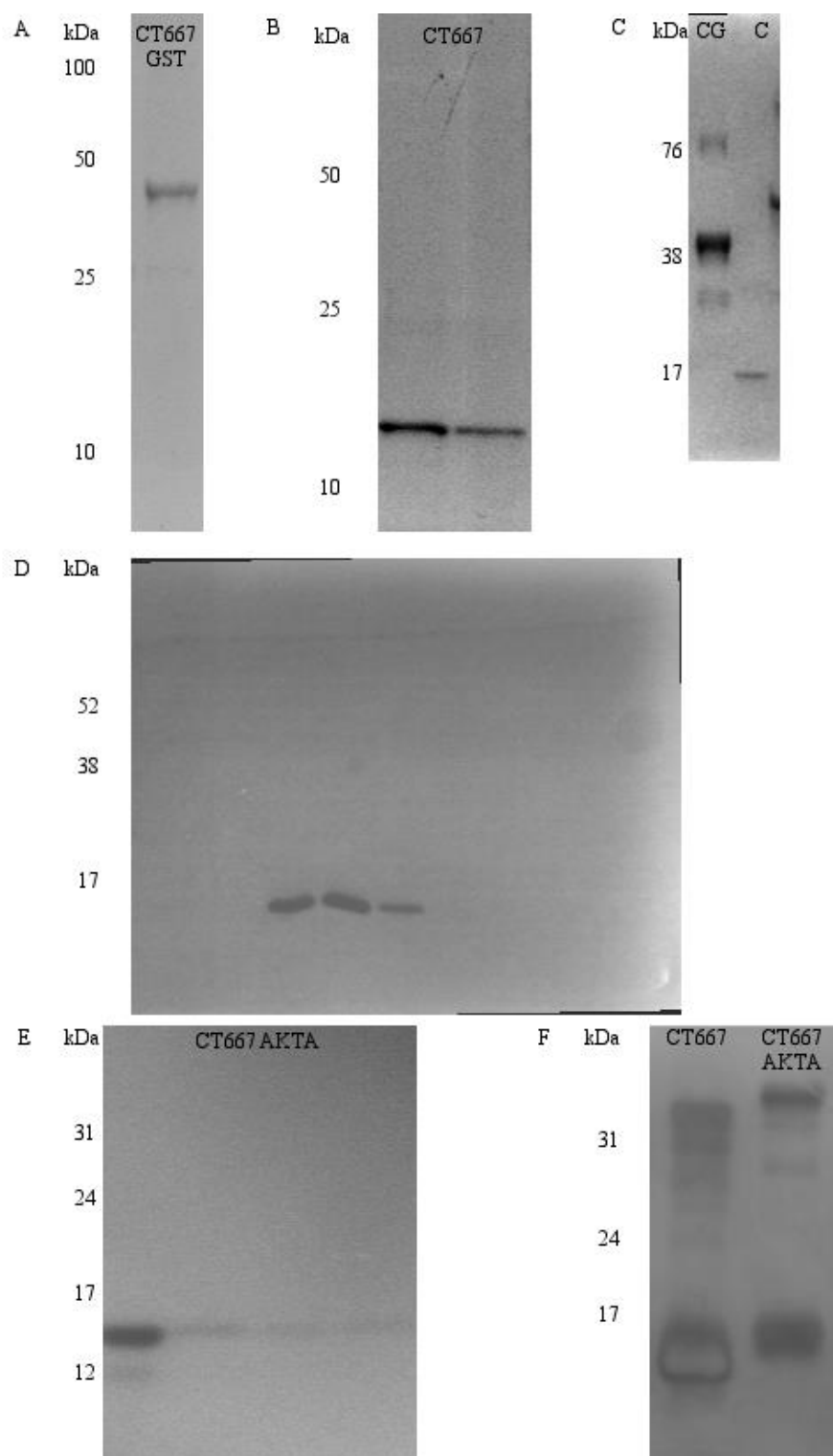


Figure 5.13. Expression trials of CT667-GST.

Western blots were incubated with an anti-GST monoclonal antibody to bind the GST-fusion moiety. The lysates were separated by ultra-centrifugation into aggregate (AGG), soluble (SOL) and insoluble fractions (INS). The relevant bands are highlighted in red. A) CT667-GST was soluble and successfully purified using the recommended conditions described previously. B) CT667 was successfully purified after cleaving the GST-fusion moiety. C) Comparison of CT667-GST and CT667. D) Silver stain of CT667 after size-exclusion based chromatography purification (CT667AKTA). E) Comparison of CT667 pre and post-purification through size-exclusion based chromatography. F) Silver stain comparison of CT667 pre and post-purification through size-exclusion based chromatography

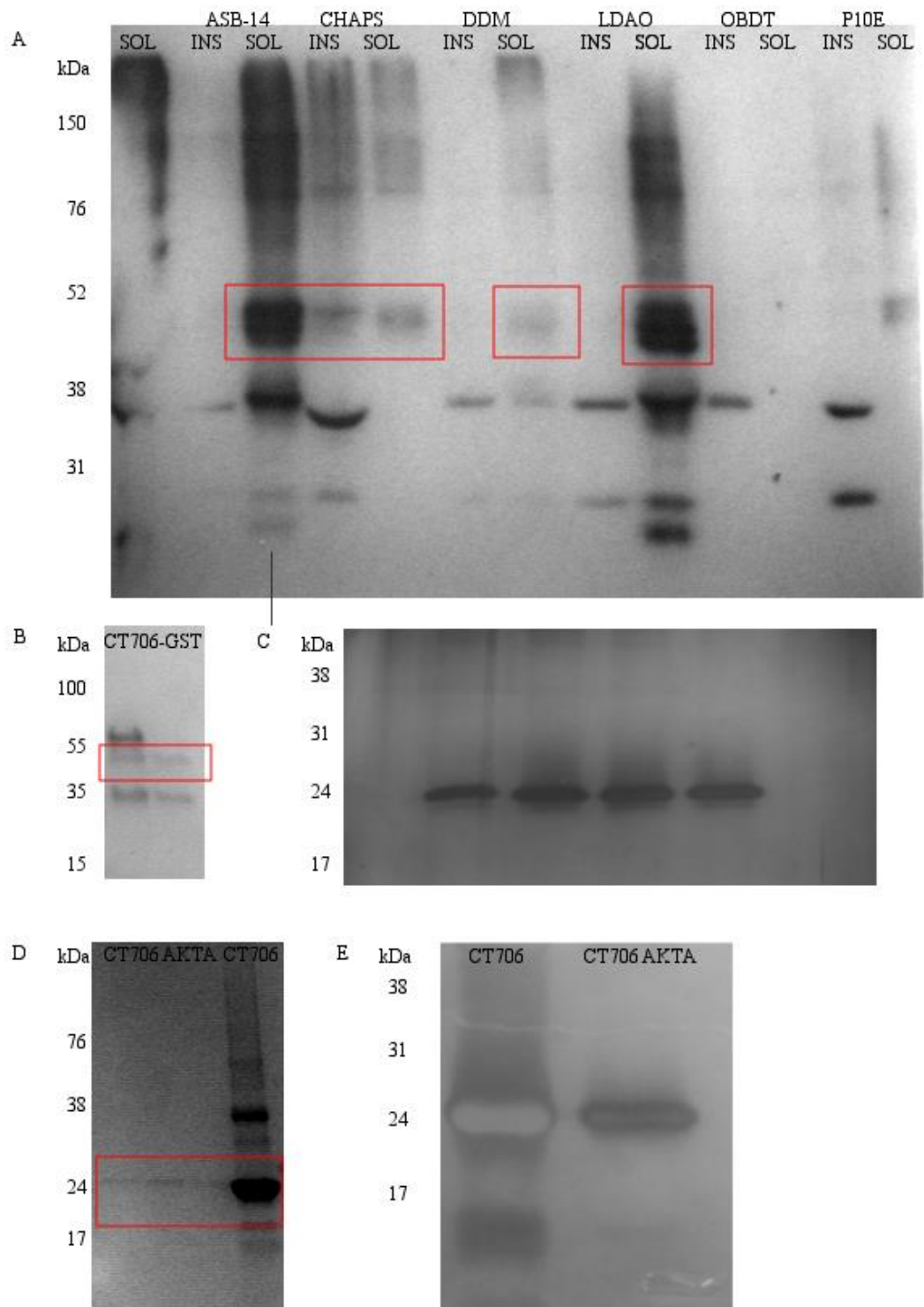


Figure 5.14. Expression trials of CT706-GST.

Western blots were incubated with an anti-GST monoclonal antibody to bind the GST-fusion moiety. The lysates were separated by ultra-centrifugation into aggregate (AGG), soluble (SOL) and insoluble fractions (INS). The relevant bands are highlighted in red. A) CT706-GST was solubilised by ASB-14, CHAPS and LDAO but not DDM, octyl- β -D-1-thioglucopyranoside (OBDT) and polyoxyethylene (10) tridecyl ether (P10E). B) CT706-GST was successfully purified after ASB-14 solubilisation. C) Silver stain of CT706 after size-exclusion based chromatography purification (CT706AKTA). D) Comparison of CT706 pre and post-purification through size-exclusion based chromatography. E) Silver stain 167 comparison of CT706 pre and post-purification through size-exclusion based chromatography.

5.2.6. ELISA validation of antibody responses associated with scarring

Initially the selected antigens were retested on the 59 sera screened with the array, excluding one sample which we no longer had serum, and the complete set of 126 sera available from the original study (Table 5.6).

Table 5.6. Patient demographics in adults with and without scarring, complete set of 116 sera.

Associations were determined using a generalised linear model.

	HEALTHY CONTROLS	SCARRED CASES	ASSOCIATED P-VALUE
NUMBER	58	58	NA
AGE IN YEARS (95% CI)	55.50 (30.43-73.73)	60.00 (34.00- 77.88)	0.199
FEMALE (N [%])	40 (68.97)	39 (67.24)	0.842

The positive control anti-Pgp3 results from the ELISA were very strongly correlated with those from the array, rho 0.91 and p-value < 0.001 (Figure 5.15). The remaining three antigens CT442, CT667 and CT706 had correlations with rho values of 0.26, 0.59 and 0.58 and p-values of 0.046, < 0.001 and < 0.001 respectively (Figures 5.16, 5.17 and 5.18). All three trended towards the same healthy or scarring-association identified from the micro-array, however only CT667 was significantly associated with scarring (p-values 0.252, 0.024 and 0.169). It is important to note the 1 sample lost with no remaining sera was one of the 3 individuals who had a strong antibody response against CT442. None of these differentially recognised antigens were associated with healthy or scarred individuals when tested on the complete set of 116 samples (Figure 5.19), with p-values of 0.368, 0.169 and 0.289 respectively. The strongest responses to CT442 and CT667 were found in healthy individuals, but no clear association with protection was found.

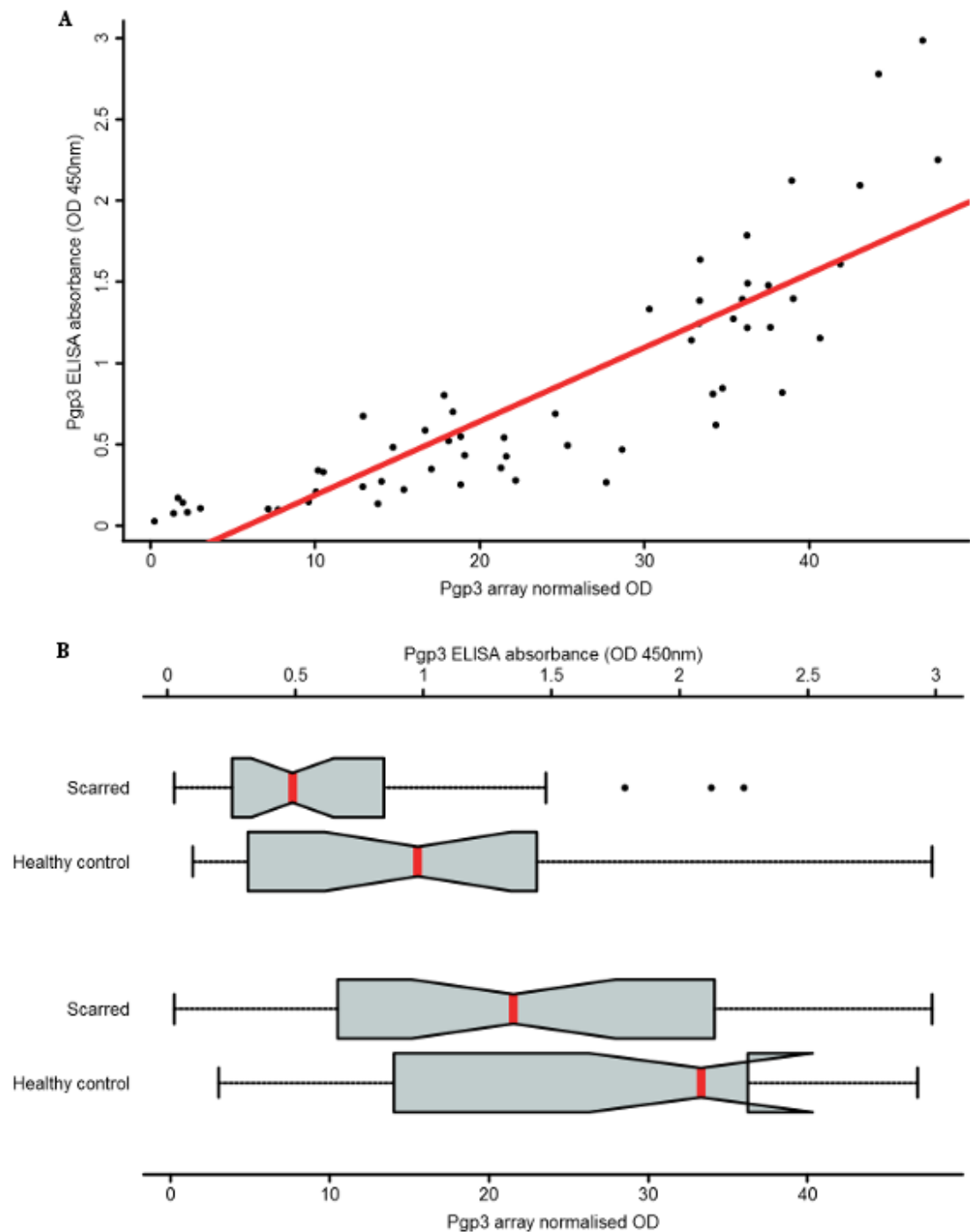


Figure 5.15. Pgp3 correlation between ELISA and array results.

Responses to Pgp3 were retested in 59 arrayed serum samples using and in-house ELISA. A) Correlation between ELISA and array results was high. B) Responses to Pgp3 were not associated with scarring. A linear model of results from the ELISA and micro titre plate array was used to fit the line (red).

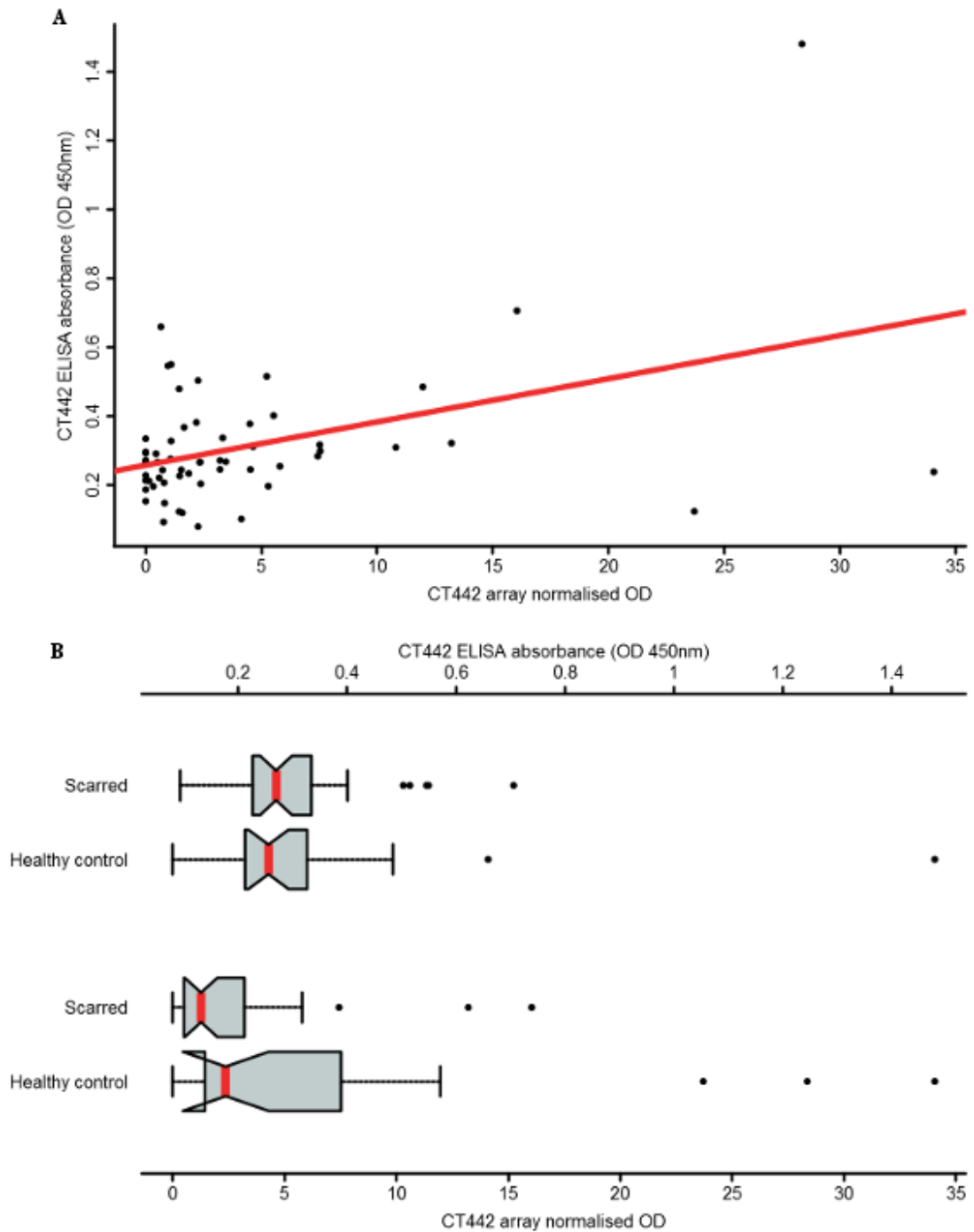


Figure 5.16. CT442 correlation between ELISA and microtitre plate array results.

Responses to CT442 were retested in 59 arrayed serum samples using an in-house ELISA. A) Correlation between ELISA and array results was poor. B) Responses to CT442 were not associated with a lack of scarring by ELISA. A linear model of results from the ELISA and array was used to fit the line (red).

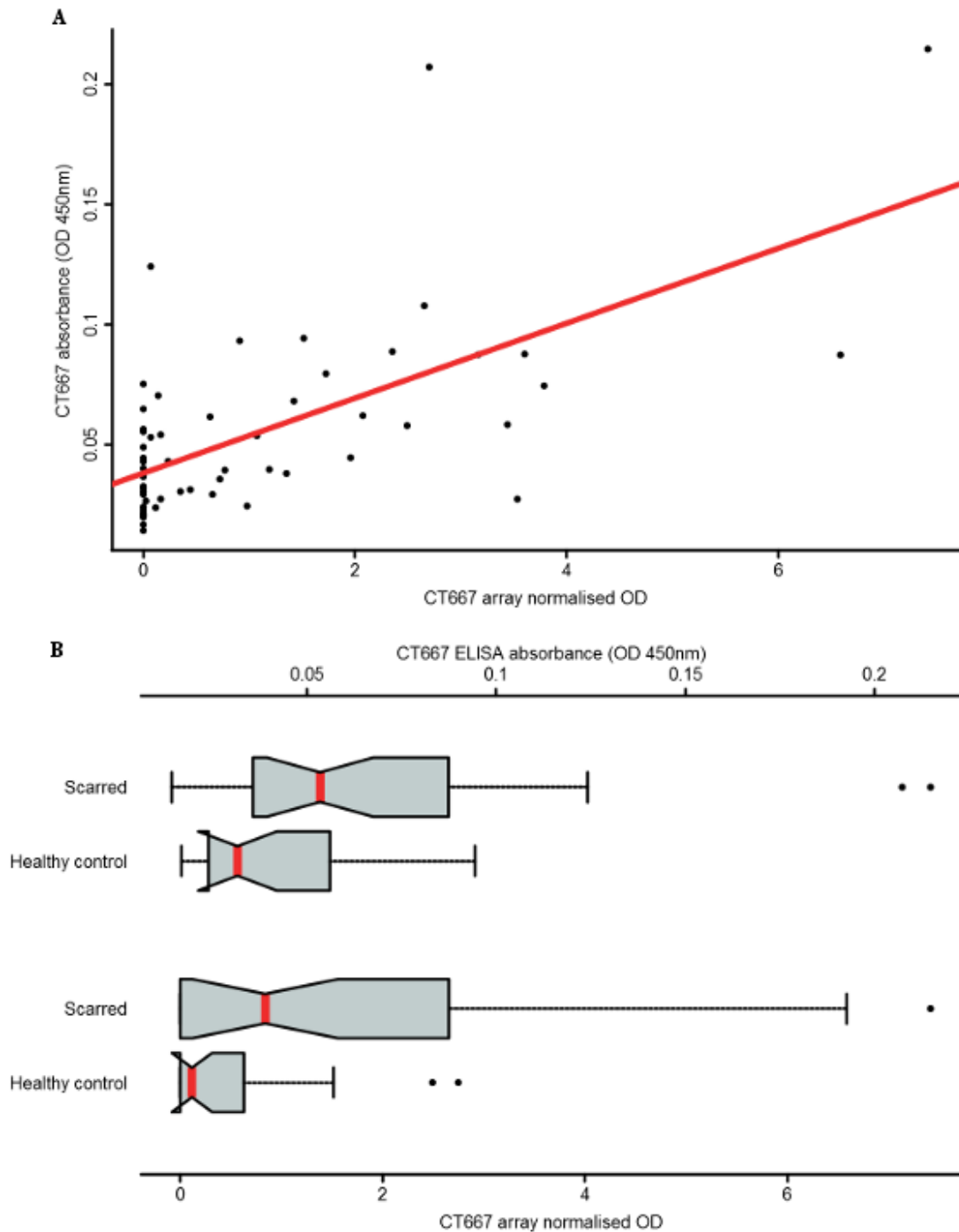


Figure 5.17. CT667 correlation between ELISA and array results.

Responses to CT667 were retested in 59 arrayed serum samples using and in-house ELISA. A) Correlation between ELISA and array results was high. B) Responses to CT667 were consistently associated with scarring by ELISA. A linear model of results from the ELISA and array was used to fit the line (red).

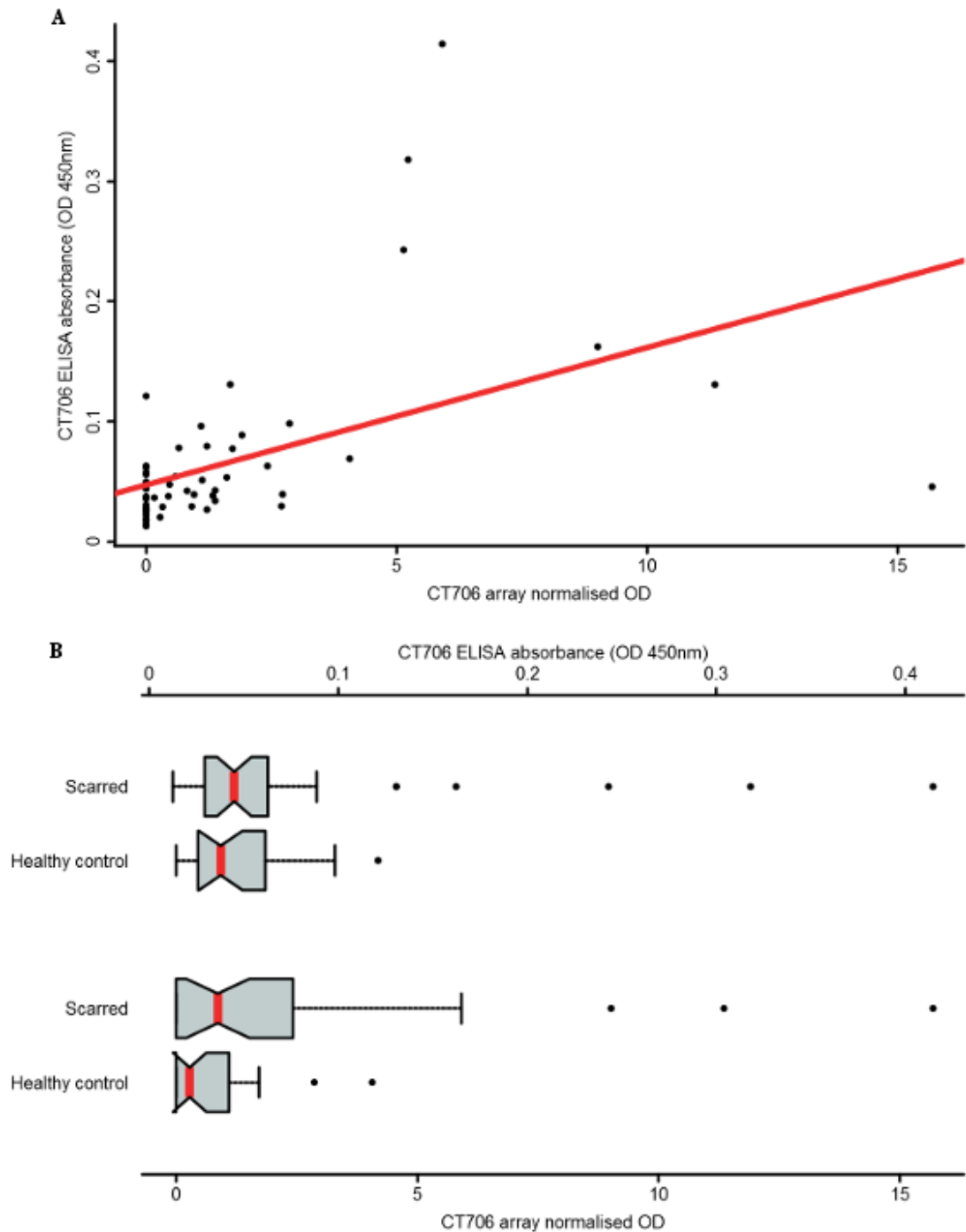


Figure 5.18. CT706 correlation between ELISA and array results.

Responses to CT706 were retested in 59 arrayed serum samples using and in-house ELISA. A) Correlation between ELISA and array results was high. B) Responses to CT706 were not associated with scarring by ELISA. A linear model of results from the ELISA and array was used to fit the line (red).

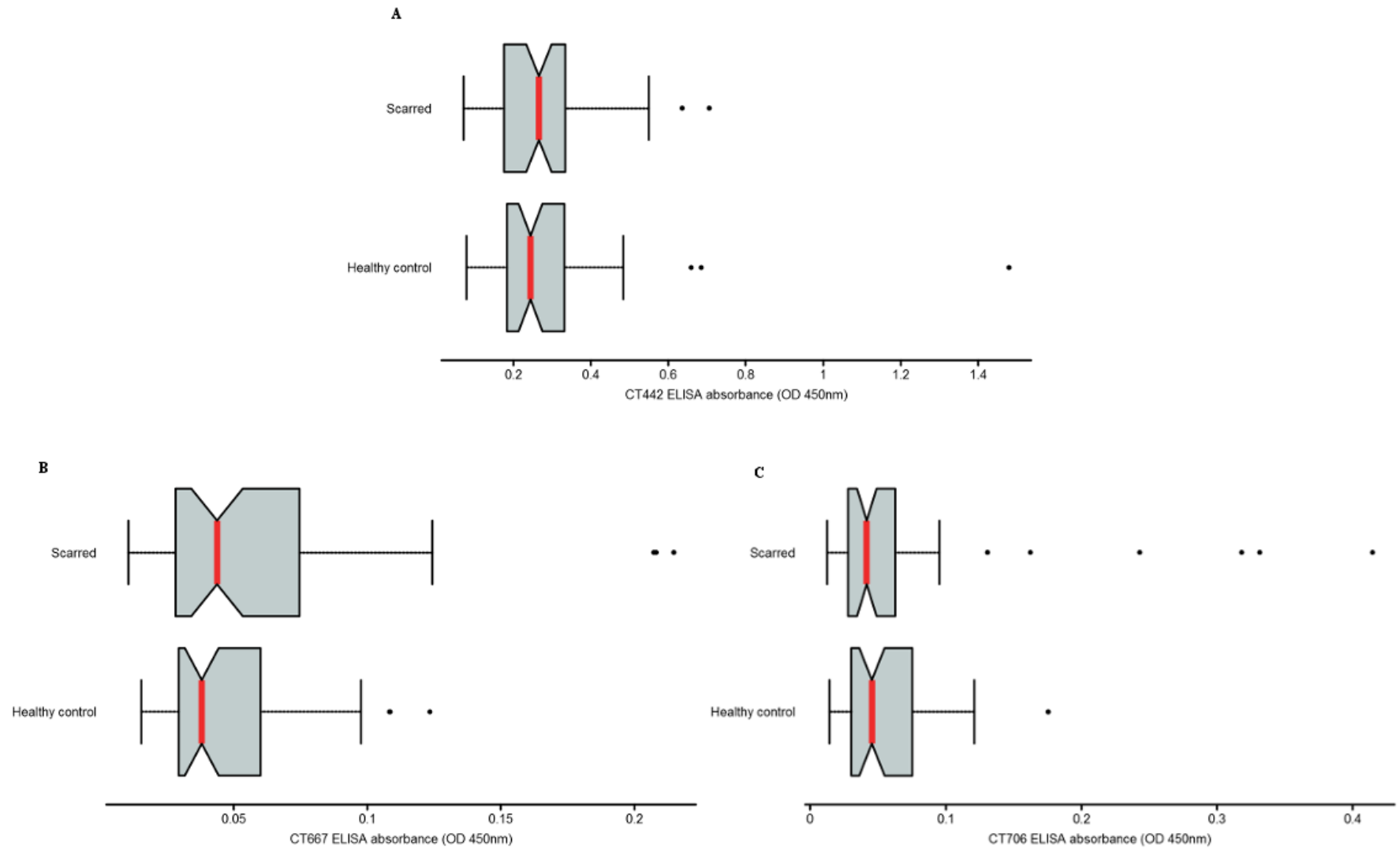


Figure 5.19. ELISA responses against CT442, CT667 and CT706 in the complete set of 116 sera.

None of the three antigens were associated with the presence or absence of scarring in adults from the complete set of 116 samples.

To test if the responses to these 3 antigens were associated with trachomatous pathology further testing in independent collections of sera from scarred cases and controls was conducted. A total of 231 individuals from a scarring case-control study collected in The Gambia in 1995 and 90 individuals from a similar study in The Gambia in 2011 were used. These studies were cross-sectional collections of cases with scarring and matched controls (chapter 3.4.1 and 3.4.5).

All 3 antigens showed mixed or minimal association with scarring in these studies (Table 5.7). CT442 was significantly associated with scarring in adults from the Gambian study from 2011, this was inconsistent with its association with healthy adults from the microtitre plate-array results.

Table 5.7. ELISA results from the complete 116 sera and two further scarring case-control studies.

The 3 arrayed antigens were tested on the complete set of 116 sera (2006), 231 samples from a previous scarring case-control study in The Gambia (1995) and a subsequent scarring case-control study in The Gambia (2011). CT442 showed no association with the absence of scarring, and CT667/CT706 were not associated with scarring.

ANTIGEN	SERA	HEALTHY MEDIAN (IQR)	SCARRED MEDIAN (IQR)	P- VALUE	OR (95% CI)
CT442	2006	0.24 (0.19-0.33)	0.27 (0.18-0.33)	0.368	0.33 (0.02-3.36)
	1995	0.29 (0.18-0.43)	0.30 (0.15-0.46)	0.990	1.00 (0.46-2.13)
	2011	0.31 (0.22-0.38)	0.36 (0.24-0.48)	0.023	36.19 (2.20-1109.21)
CT667	2006	0.04 (0.03-0.06)	0.04 (0.03-0.07)	0.169	2766.28 (0.06-6.20e ⁸)
	1995	0.14 (0.05-0.23)	0.12 (0.03-0.23)	0.344	0.59 (0.18-1.68)
	2011	0.03	0.03	0.802	0.23

		(0.02-0.06)	(0.02-0.07)		(2.47e ⁻⁶ - 24634.90)
CT706	2006	0.05 (0.03-0.07)	0.04 (0.03-0.06)	0.289	42.59 (0.07-1.26e ⁵)
	1995	0.14 (0.06-0.23)	0.12 (0.04-0.23)	0.307	0.58 (0.18-1.57)
	2011	0.02 (0.02-0.06)	0.03 (0.02-0.08)	0.940	1.24 (0.005-619.41)

5.2.7. Progression of scarring in a longitudinal cohort

Serum was obtained from a total of 311 individuals from a longitudinal cohort study in Moshi district, Northern Tanzania collected between 2012 to 2016 (chapter 3.4.6 and Table 5.8). This was a four-year study investigating factors related to scarring progression. All children aged 6-10 years were recruited from 3 neighbouring villages, 616 children were enrolled of which 506 were seen at the first time point. At baseline and every 3 months, conjunctival swabs were taken for Ct infection testing alongside clinical grading for trachoma. Finger prick bloodspots were taken at the 4-year time point.

Table 5.8. Demographics of children sampled at the end of a 4-year longitudinal cohort of scarring progression in Tanzania.

Associations were determined using a generalised linear model for continuous data and a X² test for categorical data.

	NO PROGRESSION	PROGRESSION	P-VALUE
NUMBER	243	68	NA
AGE IN YEARS (95% CI)	11.00 (9.00-16.00)	11.00 (9.00-15.33)	0.465
FEMALE (N [%])	139 (57.20)	46 (67.65)	0.123
BASELINE CT⁺ (N [%])	28 (13.53)	12 (20.69)	0.181

VISITS WITH F SCORE > 1 (N [%])	0.027	
0	116 (47.74)	24 (35.29)
1-2	64 (26.34)	15 (22.09)
> 2	63 (25.92)	29 (45.31)
VISITS WITH P SCORE > 1 (N [%])	< 0.001	
0	165 (67.90)	20 (29.41)
1-2	55 (22.63)	21 (30.89)
> 2	23 (9.47)	27 (39.71)
BASELINE SCARRING GRADE (N [%])	< 0.001	
0	171 (70.37)	25 (36.76)
1	21 (8.64)	15 (22.06)
2	4 (1.65)	7 (10.29)
3	11 (4.53)	11 (16.18)
FINAL SCARRING GRADE (N [%])	< 0.001	
0	219 (90.12)	0 (0.00)
1	12 (4.94)	17 (25.00)
2	4 (1.65)	13 (19.12)
3	8 (4.94)	38 (55.88)

Higher responses to CT442 were associated with a lack of scarring progression from baseline to four years in the Tanzanian cohort (Figure 5.20 and Table 5.9). This association was independent of baseline ocular Ct infection but not persistent active disease, p-values 0.047 and 0.120. CT667 antibody responses were not significantly associated with progression (Table 5.9). CT706 antibody responses were higher in individuals whose scarring progressed from baseline to four years (Figure 5.21 and Table 5.9). This was independent of persistent inflammation but not baseline Ct infection, p-values 0.010 and 0.056. The positive control Pgp3 was not associated with progression, p-value 0.988.

Table 5.9. ELISA results from the 311 children followed as part of a 4-year longitudinal study of scarring progression in Tanzania.

ANTIGEN	NO PROGRESSION MEDIAN (IQR)	PROGRESSION MEDIAN (IQR)	P-VALUE	OR (95% CI)
CT442	0.53 (0.45-0.65)	0.46 (0.40-0.58)	0.035	0.19 (0.03-0.75)
CT667	0.07 (0.6-0.09)	0.08 (0.06-0.10)	0.304	0.07 (0.00-4.86)
CT706	0.12 (0.09-0.17)	0.14 (0.10-0.19)	0.030	43.50 (1.56-1558.37)

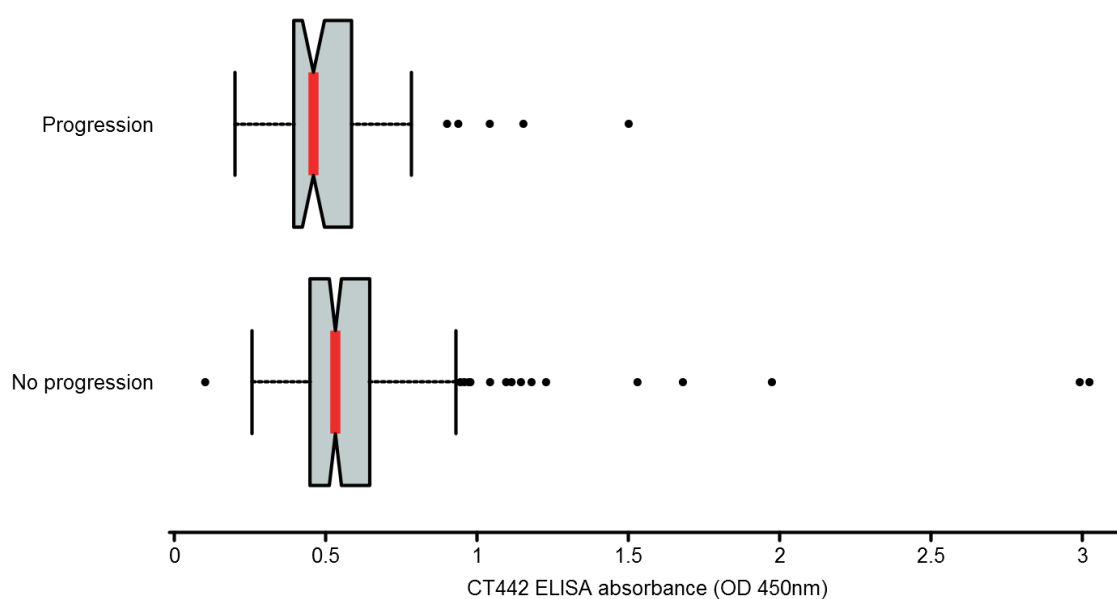


Figure 5.20. CT442 antibody responses were significantly higher in children whom conjunctival scarring had not progressed.

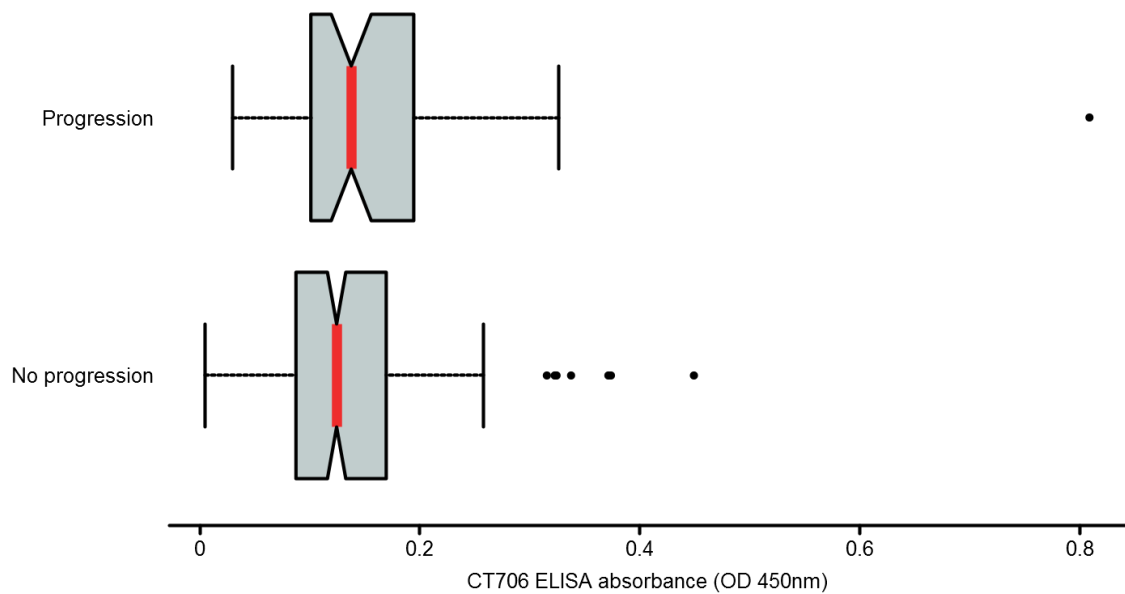


Figure 5.21. CT706 antibody responses were significantly higher in children whom conjunctival scarring had progressed.

5.3. Discussion

In this study clinical identification of scarring trachoma in a case-control study of adults from a trachoma-endemic community in The Gambia was utilised to examine the relationship between antibody responses to Ct antigens and trachomatous scarring. Antibody responses were focussed on extracellular, outer membrane and periplasmic proteins. Global antibody responses determined through screening of the Ct proteome were less informative than was found previously where responses to Ct infection were used, however there was still a trend towards more focussed responses in adults without conjunctival scarring. Heightened responses to 8 antigens were associated with the presence of scarring, only antibody responses to CT442 were associated with a lack of scarring. Independent validation of three of these antigens was achieved through further independent serological testing. CT667 and CT706 were not associated with scarring in adults in independent scarring case-control studies from the Gambia. CT706 responses were associated with progression of scarring in children during a 4-year longitudinal study in Tanzania. However, unlike in the initial microtitre plate array studies, CT442 responses were not associated with an absence or protection from scarring in adults in

the independent studies. Nevertheless, intriguingly CT442 antibody responses were associated with protection from scarring progression in a 4-year longitudinal cohort of Tanzanian children.

5.3.1. Adults have more focussed antibody responses

The number of immunogenic antigens, those which passed initial filtering, and the number of commonly recognised antigens, determined by breadth, were both lower in this study of adults than in children. This reflects the reduced frequency and duration of Ct infection and active trachomatous disease in older individuals. It also demonstrates a time dependent focussing of antibody responses. These individuals have had a lifetime of exposure to Ct promoting selective activation and maintenance of high affinity antibodies. As frequency and duration of Ct infection decreases, the opportunity for restimulation of Ct-specific plasma cells is reduced. Without reactivation these plasma cells will not be maintained. This can put antigens commonly exposed to the extracellular space and B-cells at a selective advantage. Antibodies specific to these antigens are more likely to be activated during infrequent episodes of Ct infection, promoting their activation and affinity maturation. This focussing of antibody responses was reflected in the common recognition of antigens expressed at similar stages of chlamydial development or their cellular localisation in both scarred and healthy individuals.

Proteins expressed very early and late in the Ct developmental cycle were over-represented. It is during these periods that Ct in the form of infectious EBs is most exposed to the humoral immune system. Potential targets may be EB outer membrane proteins, proteins in the membrane of extrusions and a more unpredictable group of proteins exposed to the extracellular environment upon Ct lytic exit from cells. This was supported by the over-representation of chlamydial outer membrane, extracellular and periplasmic proteins in the complete list of immunogenic antigens. Antibody responses of children were not focused on these antigens, suggesting an age dependent maturation of antibody repertoire.

Diversity of antibody responses was not significantly different between adults with and without trachomatous scarring. There was a trend towards increased diversity in scarred adults by Shannon's diversity index, this was not supported by Simpson's

diversity index. This reinforces the finding that adults in general had more focussed antibody responses. This suggests that lifetime exposure in an endemic community focuses responses on a few immunodominant antigens and gradually drives out diversity of antibody responses.

5.3.2. Few antigens were differentially recognised between individuals with and without scarring trachoma

There was no global difference in antibody responses between adults with and without scarring, however a small panel of antigens was differentially recognised between these groups. Antibody responses to 8 antigens were associated with scarring and responses to a single antigen were associated with a lack of scarring. The low number of differentially recognised antigens supports a focussing and reduced heterogeneity of antibody responses in older individuals. Three of these antigens were identified as differentially recognised in the initial analysis of this array, CT442, CT667 and CT706. Of the 7 antigens previously identified that were not supported following independent analysis, 6 had been originally differentially associated by Fishers' exact test comparing the number of positive responses between adults with and without scarring. A more objective definition of positivity was used in the follow-up independent analysis leading to no significant differentially recognised antigens being identified.

Limited published information was available for these antigens. CT314, CT425 and CT442 have been identified as immunogenic in patients with urogenital Ct or related disease¹⁶⁹, mouse models have shown CT442 can also induce CD8⁺ T-cell responses¹⁹⁵. No functional studies have been undertaken of these antigens, although some of them share homology with bacterial proteins of known functions. These relate to cell division (CT471 and CT697), transcription (CT314), protein quality control (CT706) and type-3 secretion (CT667). These homologues and localisation predictions suggest they likely reside within the inclusion, meaning they would only be exposed to B-cells upon cell lysis. There is evidence that some of these targets may be more easily accessible to the host immune system. Proteomic analysis of Ct EBs identified CT314 as a component of the outer membrane complex³²⁹. CT442 has been localised to the inclusion membrane¹⁹⁵ and during early stages of the Ct developmental cycle, CT667 appears to reside in the cytosol³²⁸. CT706 is a homologue of clpP proteases, proteases

including CPAF and HtrA which were also predicted to reside within the inclusion are known to be secreted into the host cytosol and can induce antibody responses^{330, 331}.

If these antigens can induce antibody responses this poses the question, how could these be associated with the presence of conjunctival scarring. Development and progression of scarring is linked to the frequency and duration of inflammation as a result of ocular Ct infection. If antibody responses can impact scarring they likely function at earlier stages of active disease during the development of scarring and its inflammation driven progression.

CT442 is an Inc and therefore could be localised to the surface of extrusions¹⁰¹, these are EB-containing membranous compartments released from Ct-infected cells. Antibodies specific to CT442 could bind the surface of extrusions promoting destruction of these infectious bodies. Reduced survival and transmission of Ct would lessen inflammatory responses, minimising conjunctival damage related to scarring. Heightened antibody responses against CT442 could also indicate upregulated expression or increased availability of this antigen to the immune system. Expression of Incs is known to vary between serovars of Ct³³², it is therefore possible there are differences within ocular serovars and strains. In this scenario CT442 could also be inducing CD8⁺ T-cell responses which have been demonstrated in mouse models. A non-human primate model of trachoma recently identified CD8⁺ T-cells as important in protective immunity to Ct infection after vaccination with a plasmid-deficient live-attenuated Ct strain²⁰⁰. It is plausible that heightened CT442 antibody responses in adults with scarring are indicative of increased accessibility to this antigen promoting protective CD8⁺ T-cell responses.

The scarring associated antigens may be related to diversity of antibody responses in children which delay the development of partial immunity, allowing more frequent and longer episodes of Ct infection. Based on the previously described decoy hypothesis, antibodies against these antigens would therefore facilitate Ct infection by diverting responses away from protective antigens, supporting prolonged and intense inflammation of the conjunctiva. Selective reactivation of plasma cells specific for these antigens would occur through repeated exposure to Ct, due to the diversity of antigens targeted. In this hypothesis, boosting of these targets would not be seen in all adults with scarring. In support of this, strong antibody responses against these antigens were infrequent but often exclusive to adults with scarring. This would also match the finding

that responses against the three differentially antigens tested from different studies of adults with and without scarring did not support their association with disease outcome.

CT314 is localised to the outer membrane complex of Ct EBs³²⁹. It is possible that antibodies against CT314 enhance Ct infectivity or block the neutralising ability of protective antibodies, as described in chapter 4. This would prolong infections and associated inflammation possibly increasing the development of conjunctival tissue damage and scarring.

Secreted bacterial proteases function to enhance Ct survival *in vitro*, CPAF can neutralise complement and antimicrobial peptides^{333, 334}. CT706 is a conserved protease and could also be functionally important for Ct survival, as described for CT442 above antibody responses against CT706 could be indicative of increased expression or accessibility rather than being directly pathogenic.

5.3.3. CT442 and CT706 antibody responses are associated with differential scarring progression

Examination of antibody responses in children at the end of a 4-year study of scarring progression in Tanzania identified heightened responses to CT442 in non-progressors and CT706 in progressors. This supports the suggestion that antigens associated with scarring likely function at an early time point in the development of scarring and its progression.

CT442 antibody responses were heightened in children in whom conjunctival scarring did not progress during the 4-year study. This effect was independent of Ct infection at baseline but not recurrent inflammation throughout the 4 years. Children with strong antibody responses against CT442 also had low levels of inflammation, both follicular and papillary hypertrophy, during the study. CT442 antibody responses could be indirectly associated with this phenotype or they could be actively involved in reduced inflammation through control of infection and infectious load. The two hypotheses discussed earlier in which CT442 antibody responses either promote Ct clearance directly or indicate increased CT442-specific CD8⁺ T-cell responses could both explain fewer inflammatory episodes.

CT706 antibody responses were heightened in children in whom conjunctival scarring progressed during the 4-year study. This effect was independent of recurrent inflammation and only partly explained by baseline Ct infection. Children with strong antibody responses against CT706 were at greater risk of scarring progression, irrespective of their inflammatory phenotype. The function and localisation of CT706 is unknown therefore it is difficult to speculate how antibody responses against it could be involved or indicative of conjunctival tissue damage and subsequent scarring. Bacterial homologues of this protein are important in degradation of misfolded proteins which in *Mycobacterium tuberculosis* has been postulated to regulate levels of potentially toxic cellular proteins³³⁵. A homologue in *Listeria monocytogenes* can enhance survival in macrophages through phagosome exit³³⁵. It is possible antibody responses against CT706 indicate changes in expression and accessibility as discussed previously, rather than being directly important.

5.4. Conclusions and future work

Antibody responses against Ct antigens were more focussed in adults, there was considerable homogeneity between those with and without conjunctival scarring. Only 9 antigens were differentially recognised between the two groups, of which only one (CT442) was associated with a lack of scarring. Limited localisation and functional information available for these antigens means predictions of how they might be targeted by antibodies and how this could impact scarring are difficult. CT442 is an inclusion membrane protein and therefore potentially accessible to B-cells through Ct cell exit via extrusions. The importance of CT442 antibody responses was strengthened by the finding of increased responses in children whose conjunctival scarring had not progressed in the previous 4 years. CT706, identified in the micro-array as scarring associated, was more frequently recognised in children whose conjunctival scarring had progressed in the previous 4 years.

Future work should examine whether the responses identified here in children with differential progression of scarring are present in adults with and without scarring in these communities. A greater understanding of the potential functions of targets identified here should also be a priority. As CT442 antibody responses were protective the ability of these antibodies to impact Ct infectivity and survival should be examined

in vitro, specifically testing in macrophages and with peripheral blood mononuclear cells if they are able to promote phagocytosis or cytolytic responses.

6. Identification of *Chlamydia trachomatis* genes under selection using whole genome sequencing and correlation with serological immunogenicity

6.1. Natural selection, pathogen adaptation and antigenic variation

6.1.1. Natural selection as an evolutionary process

Natural selection is the evolutionary process by which allele frequencies change over time in a population dependent on whether they are advantageous or deleterious for an organism's fitness. Fitness in this context broadly refers to the variable ability of individuals within a population to survive and reproduce. Mutations which change alleles arise randomly and for the most part are neutral²⁵³, meaning they neither benefit nor harm an individual's fitness. Mutations which impact fitness are subject to selection pressures which can be classified as either balancing selection or directional selection. Balancing selection maintains two or more variant alleles at one locus³³⁶. Directional selection is where one allele is favoured over another; this can increase the frequency of advantageous alleles (positive selection) or decrease the frequency of deleterious alleles (purifying selection)³³⁷.

Host-pathogen interactions present a complex picture in terms of selection, both host and pathogen have co-evolved with pressure from one driving selection in the other. In what is often described as an evolutionary 'arms race' hosts select for traits which limit or control pathogen infectivity while pathogens simultaneously select for traits which promote their survival and transmission³³⁷. In the case of humans and their pathogens, the significantly reduced lifespan of pathogens allows adaptation on a far quicker scale than humans³³⁸. This is most clearly seen in the rapid and ongoing rise of antibiotic and drug-resistance in human and zoonotic pathogens.

The strongest single force of selection acting on pathogens is driven by host innate and immune responses. Pathogen survival and ability to reproduce and transmit is directly dependent on the host immune response. Immune responses act to eliminate pathogens and prevent reinfection through a plethora of cellular and humoral pathways, this puts strong selection pressure on pathogen genes which can influence manipulation or evasion of host immunity³³⁸. This can cause changes in genes involved in infectivity and pathogenesis, particularly for intracellular pathogens where efficient entry and survival in cells is pivotal to fitness^{339, 340}. This immune response-driven selection most commonly manifests in variation in genes which are targeted by the host immune response.

6.1.2. Antigenic diversity as natural selection

To survive and transmit in hosts with highly evolved immune systems, pathogens have and continue to adapt to avoid mechanical clearance, immune system recognition and subsequent destruction³³⁹. This adaptation happens both within and between hosts³³⁸. Within-host mechanisms are systematic changes in proteins presented to the host through a combination of phase and antigenic variation; this involves changes in gene expression and translation, multi-copy variable gene families and recombination to express previously silent genes or create mosaic transcripts³³⁹. Between-host mechanisms primarily involve allelic polymorphism, in this scenario copy number and expression patterns do not necessarily change but multiple allelic forms of immunodominant surface antigens are present in the population. These polymorphic alleles typically have cyclical frequencies driven by host immune recognition, as alleles are recognised and targeted by the immune system they are selected against. Rare alleles increase in frequency until they are recognised by the immune system and are selected against.

Unlike within-host mechanisms, allelic polymorphism is directly driven by host immune responses in previously exposed hosts. Alleles recognised by the immune system are selected for if responses against them benefit the pathogen, for example opsonisation can improve cellular uptake of some intracellular bacteria, or selected against if responses promote pathogen destruction. Ability to reinfect previously exposed hosts is vital for pathogens to maintain a large enough susceptible population and to allow genetic exchange through recombination with other strains³³⁹. Polymorphic alleles under selection can be detected by searching for ‘signatures’ of selection in pathogen sequences³³⁷.

6.1.3. Markers of selection in human pathogens

The advent of whole-genome sequencing has allowed identification of regions which show local changes in variation not consistent with the rest of the genome, a number of tests have been developed to test for evidence of non-neutral mutations and predict genes under selection. These tests mostly fit into two categories; those which examine variation in allele frequency such as Tajima’s D^{253} , Fu and Li’s D and F^{341} and Fay and

Wu's H^{255} , and those which compare haplotype lengths such as integrated haplotype scores (iHS)²⁵⁶. The former compare the prevalence of low, intermediate and high frequency alleles to detect if selection is acting. In combination, they can determine whether this is positive, purifying or balancing selection (Table 6.1). The latter compare haplotype lengths around polymorphic sites between ancestral and derived sequences where changes in length indicate selection in action. The methods underlying these metrics are detailed in chapter 3.6.

Table 6.1: Evidence of natural selection by combining Tajima's D, Fu and Li's D*/F* and Fay and Wu's H

Positive and negative indicated values significantly different from zero, determined using values at every position on the chromosome. 'Any' indicates all possible values. '~ zero' indicates values not significantly different from zero. Drift is defined as no evidence of natural selection.

TYPE OF SELECTION	TAJIMA'S D	FU AND LI'S D*/F*	FAY AND WU'S H
POSITIVE	NEGATIVE	NEGATIVE	NEGATIVE
PURIFYING	NEGATIVE	NEGATIVE	POSITIVE
BALANCING	POSITIVE	POSITIVE	POSITIVE/NEGATIVE
DRIFT	ANY	~ ZERO	~ ZERO

Genome-wide analyses of selection have been conducted on a number of human pathogens and they have invariably identified host immune targets as frequently under selection. *Plasmodium falciparum* the most common causative parasite of malaria is perhaps the best studied in this respect, several studies examining isolates from Africa and South East Asia have identified genes under varying forms of selection³⁴²⁻³⁴⁶. Initial studies examined candidate genes believed to be important in host immune responses. Conway *et al* sequenced the variable surface antigen MSP1 and a selection of putatively neutral loci from 547 African *P. falciparum* isolates, they identified the block 2 region as having significant within population diversity and demonstrated that antibody responses against the most common block 2 genotype in a Gambian population were associated with reduced risk of malaria in the following six months³⁴². Related studies focussing on other *P. falciparum* surface antigens AMA1 and MSP3 have shown them

to be under balancing selection within malaria-endemic populations^{347, 348}. More recent genome-wide scans for selection have validated these results with AMA1, MSP1, MSP3 and other surface antigens under balancing selection^{343, 346}. Membrane proteins and genes expressed at the extracellular merozoite stage are overrepresented in genes that are under balancing and positive selection, supporting the influence of immune response-driven pressure.

Genome-wide studies in pathogenic bacteria including *Helicobacter pylori*³⁴⁹, *Staphylococcus aureus*³⁵⁰ and *Streptococcus pneumoniae*³⁵¹ have similarly highlighted cellular and humoral immune targets, cell surface proteins and known host-interactors as overrepresented in genes under both balancing and positive selection. This finding was further vindicated by a comprehensive study using 184 population datasets from 84 bacterial species³⁵². While the majority of each genome contained an abundance of rare nonsynonymous mutations gradually being driven out by purifying selection, as predicted for ecologically restricted organisms with small populations³⁴⁰, surface-exposed proteins consistently had evidence of balancing selection indicated by a lack of low frequency mutations. Studies in HIV-1 showed positive selection to be focussed in B-cell and CD4⁺ T-cell epitopes³⁵³, while evidence of positive selection in the two key surface glycoproteins of H1N1 influenza was similarly found primarily in B-cell and T-cell epitopes³⁵⁴.

6.1.4. Genes under selection in *Chlamydia trachomatis*

Genome-wide studies of the kind described above have been sparse in Ct. Until recently the number of sequenced genomes has been low and they have been collected from ecologically disparate sites over a timespan of approximately 50 years. There have been few if any populations of Ct isolated which could truly reflect the impact of selection within communities which have distinct evolutionary driving forces, particularly within ocular Ct. Studies comparing across the species and within variant tropisms where possible have identified some common targets of selection.

The first genome-wide study of this kind looked for evidence between two LGV isolates and either A/Har-13, an ocular isolate collected from Egypt in 1958, or D/UW3, a urogenital isolate collected from the USA in 1965¹¹⁴. Fifty-five and 63 genes had evidence of positive selection in the respective comparisons. Common to these were

known host-interactors which are either surface-exposed or secreted into the host cytosol such as; inclusion membrane proteins (Incs) D, F and G, polymorphic membrane proteins (Pmps) B, C and H and translocated actin-recruiting phosphoprotein (TARP). Two studies from Joseph *et al* in 2011¹¹⁵ and 2012³⁵⁵ utilising 12 and 32 genomes respectively looked within the Ct biovars for evidence of selection, identifying 18 and 49 ocular genes under positive selection. Immune targets and virulence factors were again overrepresented, particularly the Pmps and Incs, with the major outer membrane protein (MOMP) and heat shock protein 60 (HSP60) also highlighted. Similar results were found in a more recent study of 59 Ct isolates where they tested for selection related to ability to infect the conjunctiva and selection responsible for differences in ocular virulence³⁵⁶. Selection in 9 Pmps¹¹³ and 48 Incs³³² was independently tested using 19 and 51 Ct isolates respectively. Ten Incs and PmpB had significant evidence of positive selection, supporting these immunogenic virulence factors as being important in Ct infectivity and survival.

These studies invariably used between four and eight ocular Ct isolates from Egypt, The Gambia, Taiwan, Tanzania and Saudi Arabia which are historical isolates. These have been cultured extensively which may have introduced mutations not seen *in vivo*^{357, 358}. The genes they highlighted are likely important in differential tropism but it is unclear whether they reflect genes under selection within populations of Ct occupying the same ecological niche of the eye in trachoma-endemic communities. Such a population of Ct isolates was collected from Rombo District, northern Tanzania in the early 2000's¹⁷. These isolates were cultured prior to sequencing, as described above this appears to introduce changes not seen naturally due to selection pressures introduced by culture conditions. This chapter utilises isolates collected and direct-sequenced from conjunctival swabs as part of a study by Anna Last in the Bijagos Archipelago, Guinea-Bissau^{24, 240}. Prior to these studies, no such population was available and the only related studies focussed on *ompA*, which encodes for the immunodominant MOMP. Two populations of Ct isolates from trachoma-endemic communities in The Gambia and Tanzania had their *ompA* gene sequenced to search for polymorphisms and evidence of selection. The Tanzanian sequences were 63 % serovar A and 36 % B/Ba, they had dN/dS of 0.57 and 1.71 respectively indicating purifying and positive selection within these serovars¹⁰⁹. The Gambian sequences were 95 % serovar A and 5 % B with evidence of directional selection¹⁰⁸. A lack of balancing selection in this immunodominant antigen, which is a target of neutralising antibodies, is surprising in

the context of results from other pathogens. Genotypes were unchanged for 59 and 78 % of individuals who were Ct-positive in follow-up at 6 years (Tanzania) and two months (The Gambia) respectively. In Tanzania polymorphic *ompA* sequences were associated with a lack of clearance, while in The Gambia the serovar A genotype which accounted for 74 and 90 % of all sequences at baseline and follow-up was associated with lower load and less active clinical disease.

This homogeneity of *ompA* genotypes both with and without treatment over months and years, and mixed evidence of purifying and positive selection highlight the need for genuine population-based studies of genes under selection in ocular Ct isolates in order to better understand the interactions between Ct and the host immune system. Therefore 126 ocular Ct whole genome sequences obtained from discrete trachoma-endemic communities in the Bijagos Archipelago collected over the period of a single survey (i.e. representing contemporary strains in current circulation) were used in tests of population genetic selection to determine which genes or regions of the genome are under selection. This study aimed to provide an integrated overview of genomic identification of genes under selection with those immune targets identified through screening humoral responses across the proteome from chapters 4 and 5. Previous data from a variety of other pathogens have shown host immune pressure is a key driving force in pathogen adaptation. This analysis was combined with B-cell epitope predictions determined using bioinformatics tools, a well-tested system for identifying immunogenic regions in protein sequences³⁵⁹⁻³⁶¹. Regions that were immunogenic were compared with those under selection to both examine known targets, identify potential new targets and to narrow or focus on potential clinically important immune targets of antibody based immunity.

6.2. Results

6.2.1. Prediction of immunogenicity

Potential B-cell epitopes in our previously identified immune targets from chapters 4 and 5 were predicted using a combination of bioinformatics tools which utilise defined epitopes and structure-based predictions of immunogenicity (Table 6.2). Epitopes were identified in 35/55 targets from chapters 4 and 5. The majority of those with no

predicted epitopes were expected to localise within the inclusion, suggesting they are unlikely to be frequently exposed to the host immune system and therefore may be artefacts of the microarray. Twenty of these 35 also had CD4⁺ or CD8⁺ T-cell epitopes predicted to overlap B-cell epitopes. A literature search was performed using PubMed to identify studies where the localisation of these proteins had been determined empirically, either through proteomics of Ct infected cells or through *in vitro* studies using protein-specific antibodies. Empirically defined localisation of these proteins showed mixed agreement with software predictions, a number of those known to be secreted or reside in the outer membrane were incorrectly classified as remaining inside the inclusion. Inclusion membrane proteins (Incs) were commonly misclassified, however they are likely secreted by Ct prior to insertion into the inclusion membrane by an as yet undefined mechanism which may explain the difficulties in accurately predicting localisation.

Table 6.2. Predictions of immunogenic epitopes, transmembrane domains and cellular localisation of immune targets from chapters 4 and 5

B-cell epitopes were defined using ABCpred, BepiPred and IEDB antibody epitope predictions (chapter 3.3). T-cell epitopes were defined using SYFPEITHI, NetMHC, IEDB-SMM and ProPred (chapter 3.3). Transmembrane domains were defined using MembrainD, Phobius and TMHMM (chapter 3.3) Predicted localisations were defined using LocTree, Cello and psortB (chapter 3.3).

ID	B- CELL	T/B- CELL	MEMBRANE DOMAINS	PREDICTED LOCALISATION	EXPERIMENTAL LOCALISATION
CT017	1	1	0	Cytoplasmic	Outer membrane
CT021	0	0	0	Cytoplasmic	
CT023	5	1	0	Secreted	
CT051	6	1	0	Cytoplasmic	Inclusion lumen
CT073	3	1	0	Secreted	
CT078	2	0	0	Secreted	
CT089	10	3	0	Cytoplasmic	Secreted
CT097	5	0	0	Secreted	
CT106	3	3	0	Cytoplasmic	
CT118	1	1	2	Cytoplasmic	Inclusion membrane

CT119	2	0	2	Secreted	Inclusion membrane
CT123	2	2	0	Inner membrane	
CT142	3	0	0	Cytoplasmic	Secreted
CT168	0	0	0	Secreted	
CT181	2	1	0	Cytoplasmic	
CT223	1	0	2	Secreted	Inclusion membrane
CT228	2	0	2	Inner membrane	Inclusion membrane
CT237	1	0	0	Inner membrane	
CT284	5	0	0	Cytoplasmic	
CT316	1	0	0	Cytoplasmic	
CT381	1	0	0	Cytoplasmic	
CT494	1	0	1	Secreted	
CT502	1	0	0	Inner membrane	
CT541	3	0	0	Cytoplasmic	Outer membrane
CT545	8	5	0	Periplasmic	
CT570	1	1	3	Cytoplasmic	Periplasmic
CT579	5	2	0	Inner membrane	Outer membrane
CT584	0	0	0	Secreted	Secreted
CT592	8	1	0	Cytoplasmic	
CT642	0	0	2	Periplasmic	
CT664	9	0	1	Inner membrane	
CT668	3	0	0	Cytoplasmic	Secreted
CT694	7	4	0	Secreted	Secreted
CT695	4	0	0	NA	Secreted
CT703	3	0	0	Cytoplasmic	
CT728	1	0	4	Cytoplasmic	
CT764	0	0	1	Inner membrane	
CT795	1	0	0	Inner membrane	Secreted
CT806	8	3	0	Inner membrane	
CT813	3	0	2	Periplasmic	Inclusion membrane
CT841	5	0	2	Cytoplasmic	
CT875	3	0	0	Inner membrane	Secreted
CT029	0	0	0	Cytoplasmic	

CT334	6	1	0	Cytoplasmic
CT391	0	0	0	Periplasmic
CT630	1	0	0	Cytoplasmic
CT314	8	2	0	Cytoplasmic
CT425	8	2	0	Cytoplasmic
CT442	2	0	2	Outer membrane Inclusion membrane
CT471	0	0	1	Cytoplasmic
CT645	0	0	2	Inner membrane
CT667	1	1	0	Cytoplasmic
CT679	2	1	0	Cytoplasmic
CT698	2	0	0	Cytoplasmic
CT706	2	0	0	Cytoplasmic

6.2.2. Genome-wide evidence of purifying and positive selection, with minimal balancing selection

Sequence data from 126 ocular Ct isolates from the Bijagos Islands, Guinea-Bissau was examined for evidence of departures from non-neutral selection. These isolates were collected and sequenced as described previously (chapter 3.6)²⁴⁰. Briefly, a cross-sectional population-based survey was undertaken in trachoma-endemic communities on the Bijagós Archipelago of Guinea Bissau. Conjunctival swabs were obtained from the left upper tarsal conjunctiva of each participant. For eight individuals, whole genome sequence (WGS) data was obtained following Ct isolation in cell culture. For the remaining individuals (118), WGS data were obtained directly from clinical samples.

Tajima's D values were negative or zero for 910/953 (95.49 %) of genes, the median was -1.39 [95 % CI -2.39-0.30] (Figure 6.1). One hundred and ninety-eight genes showed significant departure from neutrality based on a minimum number of calls per site of 50 ($2.044 \leq D \leq -1.8$)²⁵³, all of which returned negative D values. This showed an excess of low frequency mutations, indicative of either population expansion or directional selection. There was no clear evidence of balancing selection.

Tajima's D values correlated strongly with Fu and Li's D* and F*²⁵⁴ (rho values 0.80 p-value < 0.001 and 0.85 p-values < 0.001 respectively). Both D* and F* had a significantly lower median, -2.78 and -2.75 respectively, suggesting an excess of singleton mutations which are found in only one isolate (Figure 6.2).

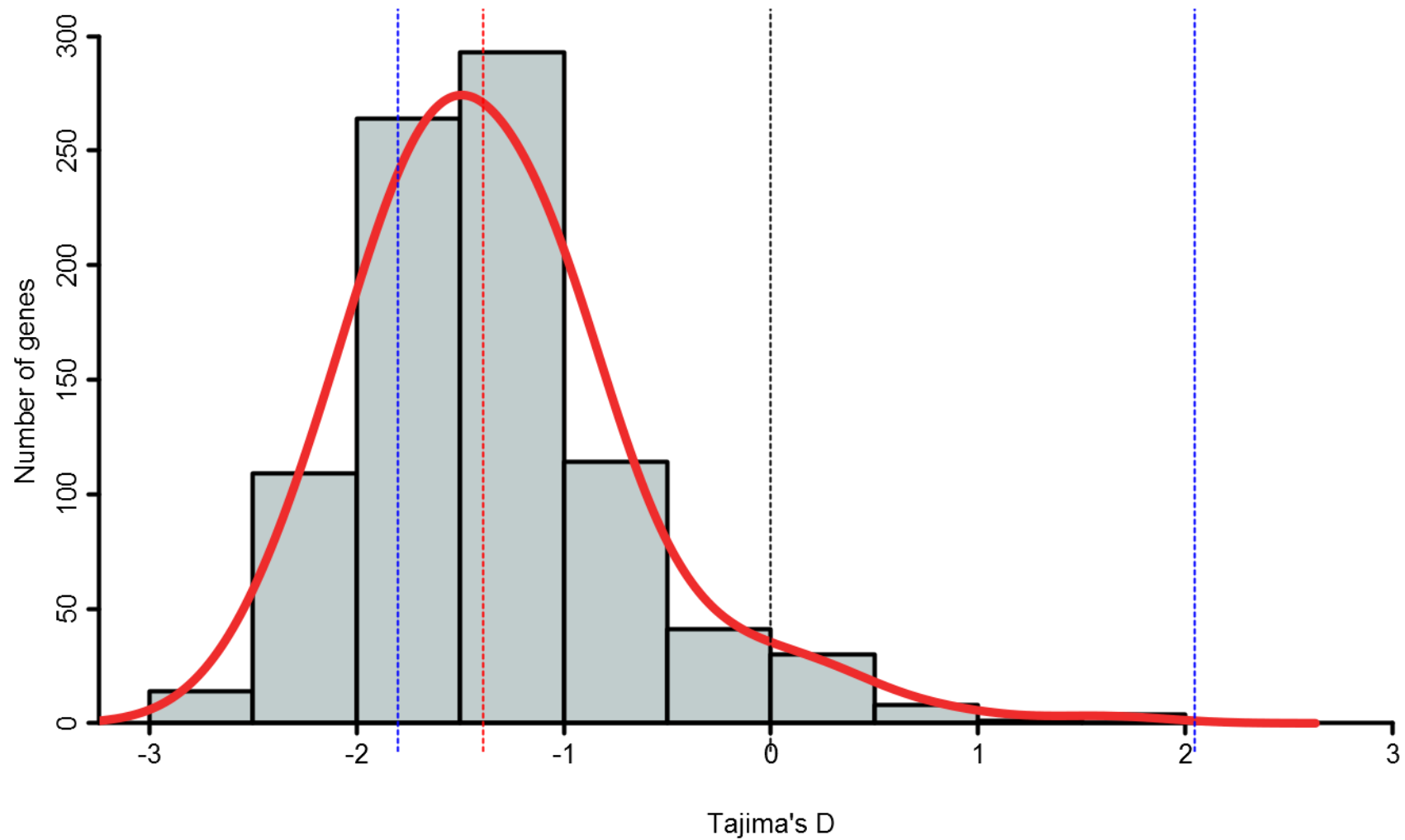


Figure 6.1. Histogram of the genome-wide distribution of Tajima's D.

The median was -1.39 (dashed red line), significant values were determined at the 0.05 level defined previously (blue lines).

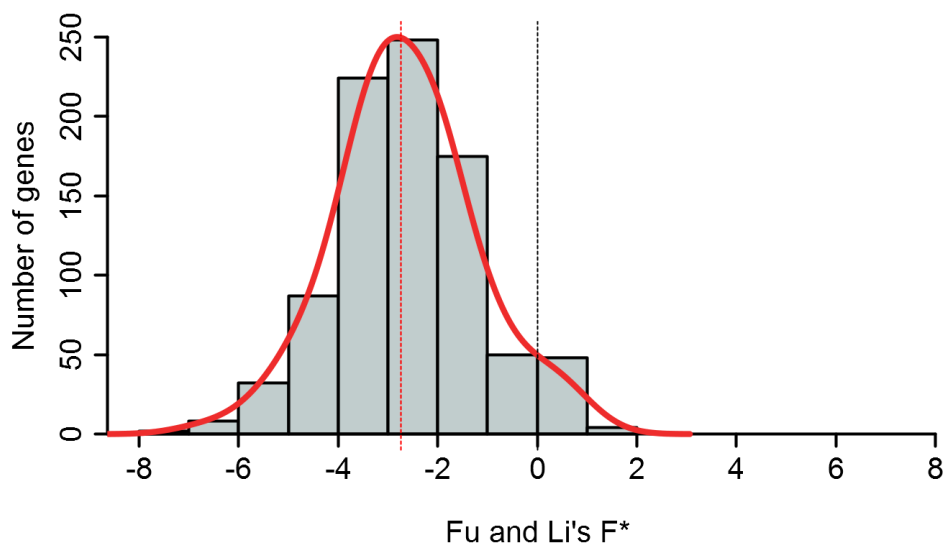
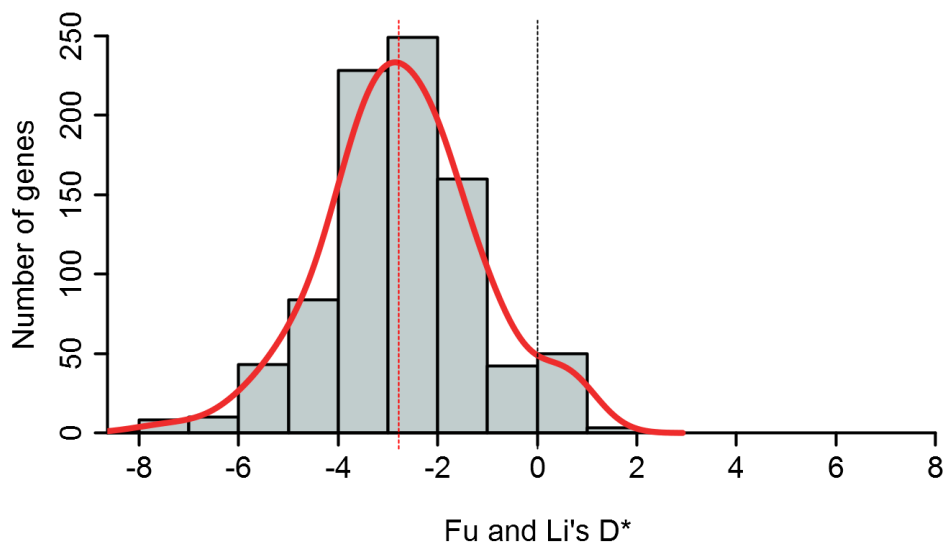
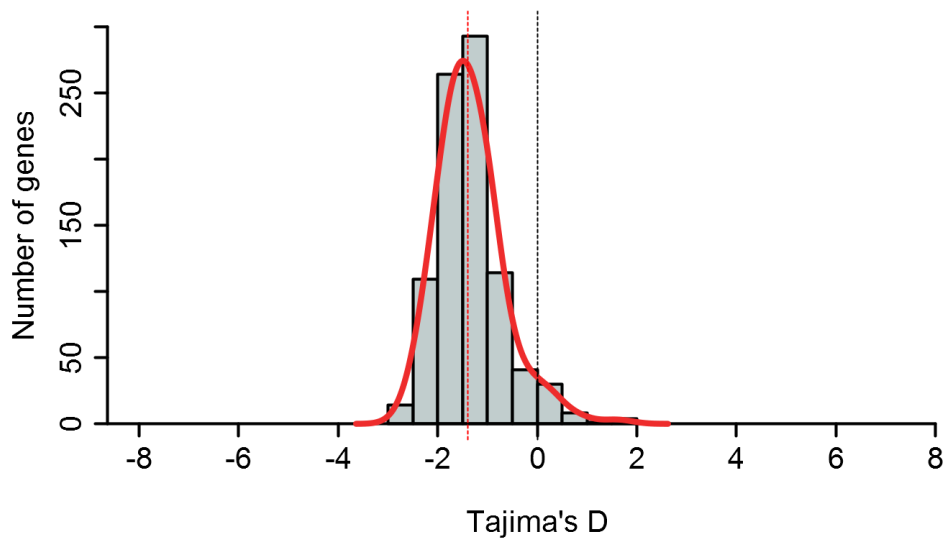


Figure 6.2. Genome-wide distributions of Tajima's D , Fu and Li's D^* and F^* .

The median values were -1.39, -2.78 and -2.75 respectively (dashed red lines).

Tajima's D values were also determined using a sliding window analysis with a window size of 42 base pairs, which equates to the common length of antibody epitopes of sixteen amino acids. Tajima's D values were negative or zero for 296059/304525 (97.22 %) windows, the median was -1.04 [95% CI -1.55-0.97]. Four hundred and twenty-one windows across 27 genes showed significant departure from neutrality, similar to gene-level results they were all negative (Figure 6.3). This supported an excess of low frequency mutations across the genome, again supporting either population expansion or directional selection. Most windows with evidence of selection were found within genes with evidence of selection at gene-level.

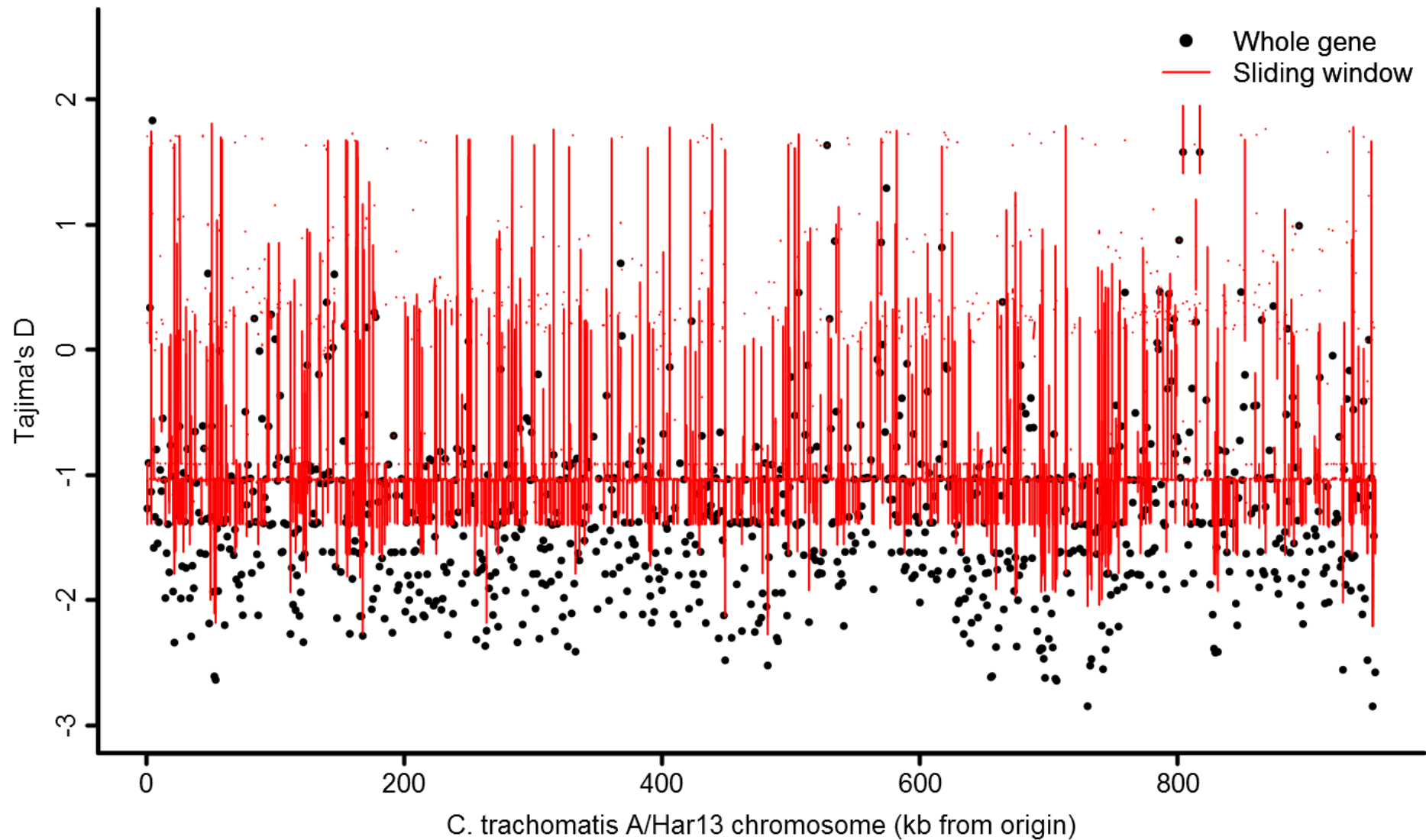


Figure 6.3. Genome-wide evidence of selection by Tajima's D.

There was clear overlap of genes with significantly negative D values (black dots) and 42 nucleotide windows with significantly negative D values (red lines). The median value of sliding windows was -1.04.

Twenty-two of 198 genes with evidence of selection were outside the 95 % distribution of the Ct genome (Table 6.3). These include a number of genes previously identified as polymorphic, notably 3 members of the polymorphic membrane protein (*pmp*) family. CT642 and CT442 to which antibody responses were previously shown to associate with susceptibility to infection and scarring progression respectively, also had evidence of selection. Secreted and outer membrane proteins were over-represented in the genes under selection, p-value 0.003 (Figure 6.4A).

All 27 genes with evidence of selection at the epitope-level were outside the 95 % distribution (Table 6.3). Twelve of these genes had gene-level evidence of selection outside the 95 % distribution, a further 11 were within the 95 % distribution but significantly different from neutrality. Secreted and outer membrane proteins were also over-represented in the genes under selection, p-value 0.003 (Figure 6.4B).

No genes or windows had evidence of balancing selection at the significance level of 0.05. To investigate genes different from the rest of the genome but not significant at this level the accepted significance level was extended to 0.1 ($1.723 \leq D \leq -1.570$). This identified a single gene and 119 windows across 18 genes which had evidence of balancing selection at the significance level of 0.1 (Table 6.4).

Table 6.3. Genes under selection identified by Tajima's D.

The top 22 genes had gene-level evidence of selection. The bottom 15 had epitope-level evidence of selection. Twelve of those with gene-level evidence of selection also had epitope-level evidence. Theta was calculated as the pairwise nucleotide diversity per site. Tajima's D SW indicates the number of sliding windows with significant evidence of selection in the respective genes.

ID	NAME	NUMBER OF SNPS	THETA	TAJIMA'S D	TAJIMA'S D SW
CT872	<i>pmpH</i>	68	0.005	-2.85	108
CT674	<i>yscC</i>	51	0.004	-2.85	34
CT652	<i>recD_2</i>	33	0.004	-2.64	4
CT050		45	0.005	-2.64	42
CT651		51	0.006	-2.63	21
CT643	<i>topA</i>	21	0.002	-2.62	3
CT604	<i>groEL_2</i>	27	0.004	-2.61	
CT049		29	0.004	-2.61	14
CT605		22	0.004	-2.61	
CT874	<i>pmpI</i>	20	0.002	-2.57	
CT852		19	0.006	-2.55	14
CT682	<i>pbpB</i>	18	0.001	-2.55	
CT675	<i>karG</i>	17	0.003	-2.52	
CT442	<i>crpA</i>	28	0.012	-2.52	30
CT414	<i>pmpC</i>	41	0.002	-2.48	14
CT868		19	0.003	-2.48	
CT676	<i>DubI</i>	17	0.006	-2.47	7
CT642		20	0.005	-2.47	
CT760	<i>ftsW</i>	17	0.003	-2.42	
CT762	<i>murC</i>	19	0.002	-2.41	3
CT311		14	0.004	-2.41	
CT639		20	0.001	-2.40	
CT641	<i>ygeD</i>	42	0.005	-2.38	2
CT640	<i>recC</i>	31	0.002	-2.38	12
CT649	<i>ygfA</i>	12	0.004	-2.38	9

CT607	<i>ung</i>	14	0.004	-2.37	1
CT623		12	0.002	-2.37	2
CT159		19	0.004	-2.28	23
CT105		45	0.006	-2.27	13
CT242		9	0.003	-2.24	13
CT681	<i>ompA</i>	25	0.004	-2.20	11
CT470	<i>recO</i>	8	0.002	-2.17	5
CT157		28	0.005	-2.13	1
CT622		47	0.005	-1.66	13
CT147		32	0.001	-1.33	1
CT046	<i>hctB</i>	39	0.013	-1.33	12
CT680	<i>rs2</i>	13	0.003	-1.12	9

Table 6.4. Genes under balancing selection by Tajima's D.

Genes were ordered by position in the Ct D/UW3 genome.

ID	NAME	NUMBER OF SNPS	THETA	TAJIMA'S D	TAJIMA'S D SW
CT003	<i>gatA</i>	6	< 0.001	-1.13	9
CT004	<i>gatB</i>	2	< 0.001	1.83	
CT047		7	0.001	-0.61	2
CT151		12	0.002	-1.83	14
CT294	<i>sodM</i>	5	0.002	-0.84	7
CT374	<i>arcD</i>	3	0.001	-0.14	1
CT404		8	0.002	-1.34	2
CT464		2	< 0.001	0.46	1
CT526	<i>r123</i>	2	0.001	1.30	14
CT529		7	0.002	-1.20	14
CT534	<i>cutE</i>	9	0.001	-0.78	4
CT542	<i>aspS</i>	11	0.001	-1.46	14
CT570	<i>gspF</i>	6	0.001	-0.75	14
CT632		10	0.001	-0.62	5
CT650	<i>recA</i>	6	0.001	-0.67	14
CT657		7	0.004	-1.27	4

CT798	<i>glgA</i>	9	0.001	-1.25	14
CTA_0934		3	0.002	-0.16	14
CT859	<i>lytB</i>	6	0.001	-0.48	3

Significant values of D can be caused by population changes as well as natural selection, to identify genes genuinely under selection and to distinguish between purifying and positive selection an outgroup was used, A/Har 13, to calculate Fay and Wu's H. The median of all H values was 0.06 [95 % CI -3.85-0.72]. This suggests the majority of negative D values were caused by population expansion rather than any force of selection. Forty-eight genes were outside the genome-wide 95 % distribution for H values (Figure 6.4). Ten genes with significantly negative H values had significantly negative D values and 9 genes with significantly positive H values had significantly negative D values, suggesting these genes are under positive and purifying selection respectively (Figure 6.5 and Table 6.5). An additional 2 genes with the lowest H values had borderline significant evidence of selection from Tajima's D and may also be under positive selection.

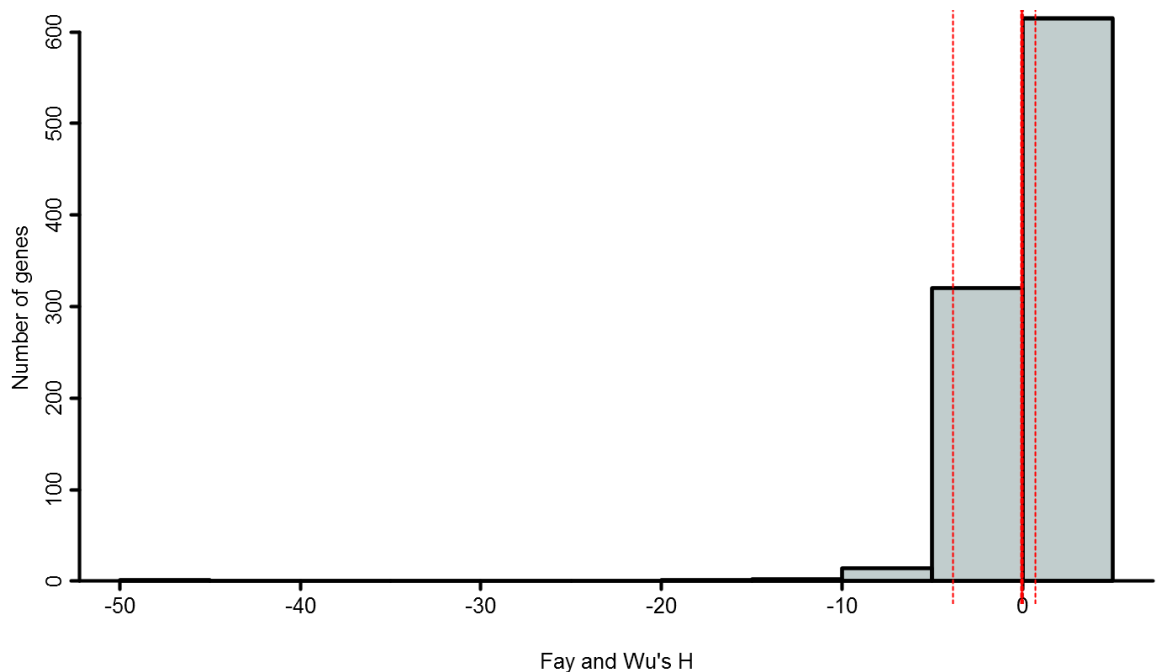


Figure 6.4. Genome-wide distribution of Fay and Wu's H.

The median value was 0.06 [95 % CI -3.85-0.72] (dashed red lines).

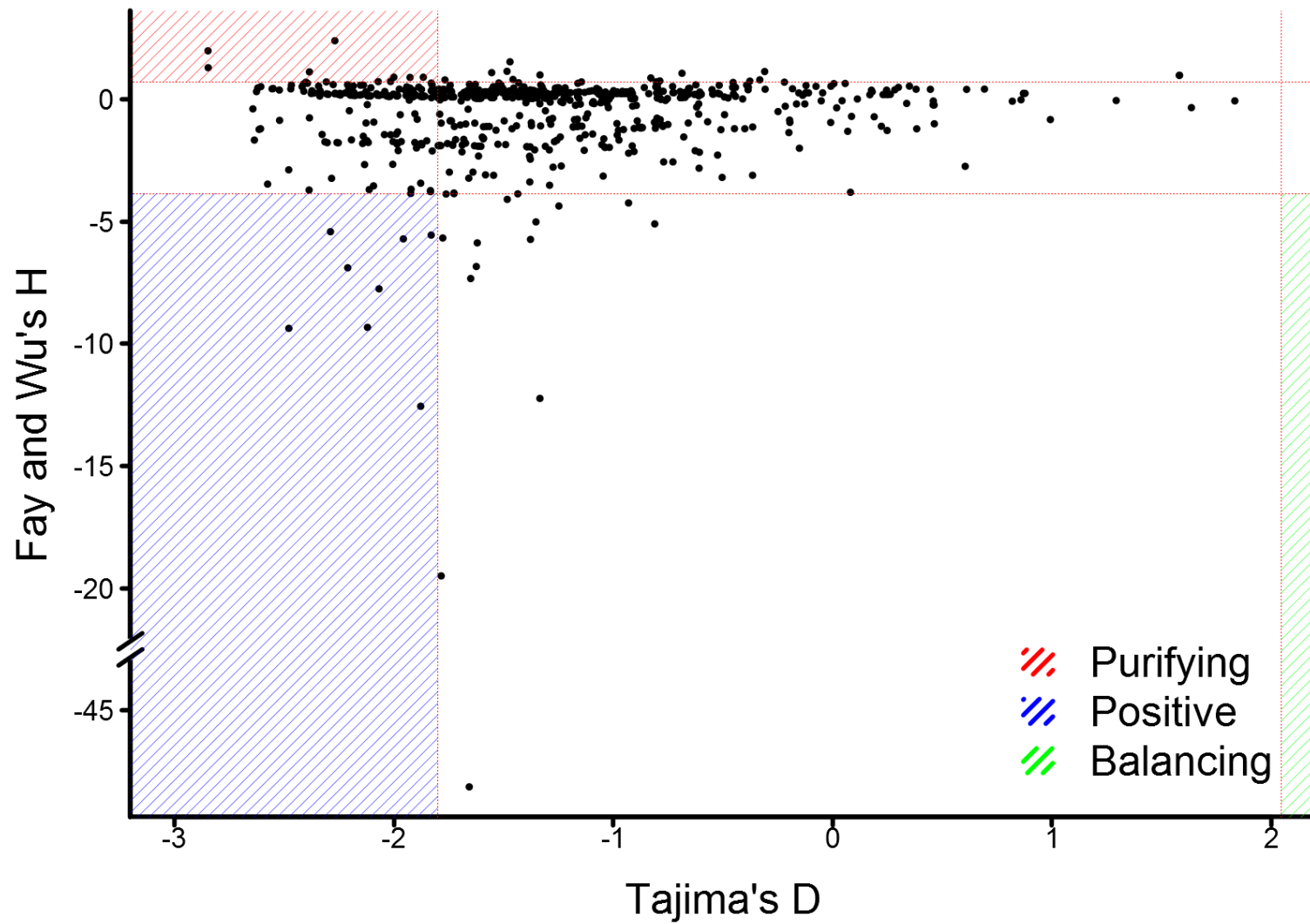


Figure 6.5. Correlation of Tajima's D and Fay and Wu's H.

Values significantly different from zero are indicated for each measure (dashed red lines). Type of selection is indicated (shaded areas).

Table 6.5. Genes under selection identified by Tajima's D and Fay and Wu's H.

The top 10 genes had evidence of positive selection. The bottom 9 genes had evidence of purifying selection.

ID	NAME	NUMBER OF SNPS	FAY AND WU'S H	TAJIMA'S D
CT033	<i>recD_1</i>	15	-5.41	-2.29
CT082		12	-9.32	-2.12
CT288		12	-12.55	-1.88
CT359		8	-5.55	-1.83
CT386		7	-7.75	-2.07
CT622		47	-48.13	-1.66
CT686	<i>sufD</i>	20	-19.50	-1.78
CT688	<i>parB</i>	7	-5.71	-1.96
CT694		14	-6.89	-2.21
CT868	<i>Dub1</i>	19	-9.36	-2.48
CT105		45	2.40	-2.27
CT223		15	0.91	-1.87
CT394	<i>hrcA</i>	15	0.73	-2.01
CT621		16	0.90	-1.93
CT624	<i>mviN</i>	17	0.91	-2.00
CT636	<i>greA</i>	15	0.73	-2.07
CT641	<i>ygeD</i>	42	1.13	-2.38
CT674	<i>yscC</i>	51	1.30	-2.85
CT872	<i>pmpH</i>	68	1.99	-2.85

Fay and Wu's H was analysed across windows of 42 nucleotides as described for Tajima's D. Windows with no mutations were automatically assigned a value of zero when determining H values, since the majority of windows had a value of zero the median was zero (95 % CI 0.00-0.11). To determine windows with significant evidence of selection the 95 % distribution was calculated excluding these zero values, the median of the non-zero windows was 0.02 (95 % CI -1.95-0.25). Six hundred and sixty-one windows across 47 genes had significantly negative values of H and 606 windows

across 58 genes had significantly positive values of H, 10 (positive) and 6 (purifying) of these had evidence of selection by Tajima's D (Figure 6.6 and Table 6.6).

Table 6.6. Genes with sliding window-level evidence of selection by Tajima's D and Fay and Wu's H.

Type of selection is indicated. Windows under selection indicates the number of sliding windows with significant evidence of selection in the respective genes.

ID	NAME	WINDOWS UNDER SELECTION	TYPE OF SELECTION
CT046	<i>hctB</i>	2	Purifying
CT082		1	Positive
CT105		13	Purifying
CT147		24	Positive
CT159		4	Positive
CT249		1	Purifying
CT359		1	Positive
CT414	<i>pmpC</i>	16	Purifying
CT442	<i>crpA</i>	17	Purifying
CT456	<i>tarP</i>	8	Positive
CT539		2	Positive
CT622		4	Positive
CT681	<i>ompA</i>	13	Purifying
CT686	<i>sufD</i>	21	Positive
CT688	<i>parB</i>	3	Positive
CT694		6	Positive

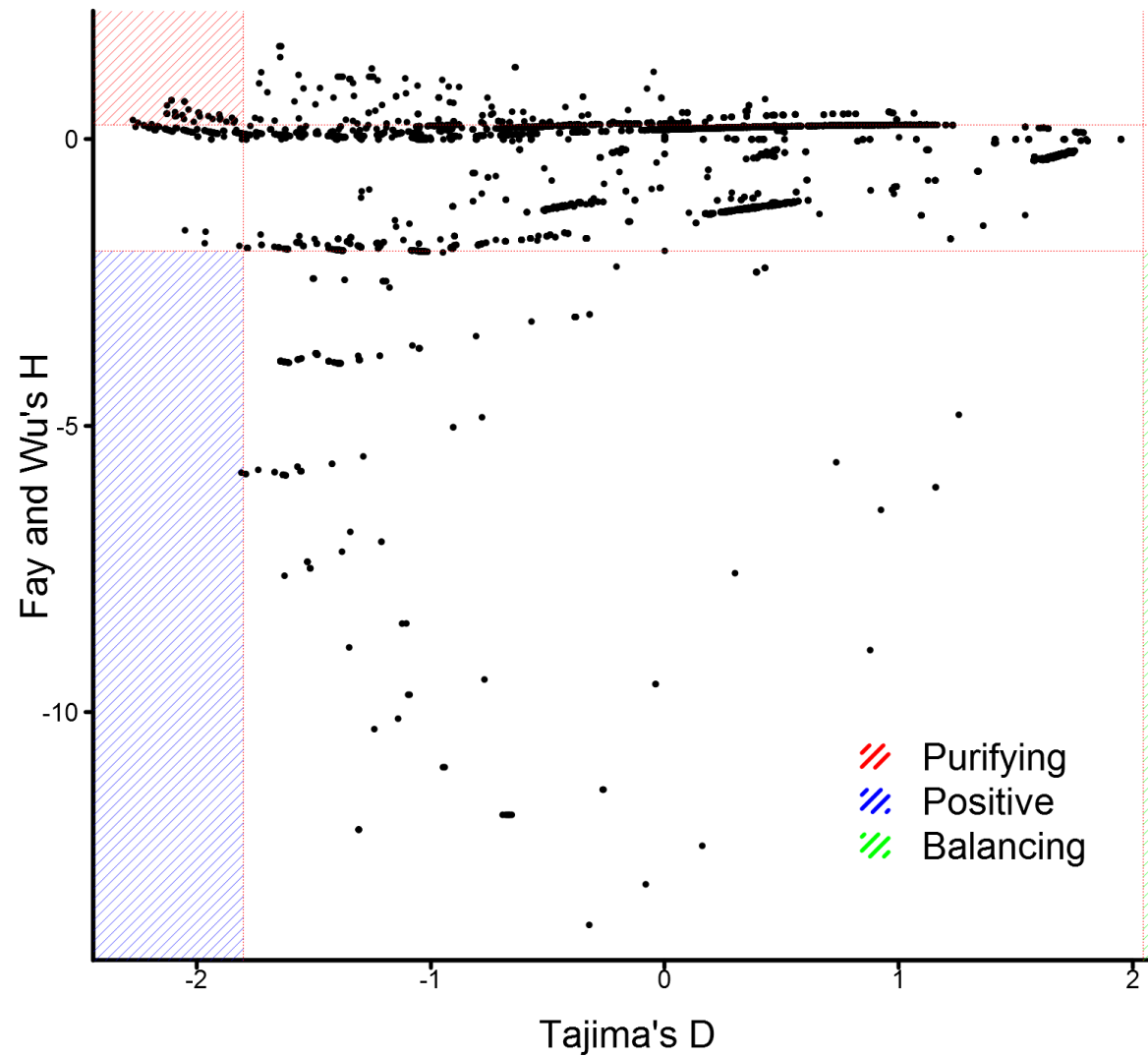


Figure 6.6. Correlation of Tajima's D and Fay and Wu's H from sliding-window analysis.

Values significantly different from zero are indicated for each measure (dashed red lines). Type of selection is indicated (shaded areas).

6.2.3. Integrated haplotype scores identify three genomic regions under positive selection

A genome-wide scan was performed to calculate integrated haplotype scores (iHS) for SNPs to validate genes under positive selection and to identify regions under positive selection. To minimise loss of SNPs missing calls were imputed using the distance-based method described in chapter 3.6. Briefly missed calls were imputed by assigning a score to each non-missing call based on the pairwise genetic distance between the isolate with the missing call and all isolates with a call. The scores were summed over the ancestral and derived alleles respectively, the allele with the lowest score was used to reclassify the missing call as it was deemed more closely related to the isolate with the missing call. Resulting iHS scores were broadly correlated with those from the raw data (Appendix Figure 4). Since there was no previous data to compare results with intrinsic biases in the results before and after imputation was used. Imputations did not significantly alter the MAF and there was no association between MAF and iHS score before or after imputation (Appendix Figure 5A). Non-imputed scores showed an inconsistent association with number of missed calls, imputed scores were higher and more variable when SNPs had greater than 15 missed calls (Appendix Figure 5B). Since there was no bias in iHS scores using imputed genotypes this data was used going forward, SNPs with greater than fifteen missed calls were excluded. The analysis included 2147 SNPs from 81 isolates, after removing sites with a minor allele frequency (MAF) less than 0.05.

iHS was determined through the rehh package in R, WHAMM, and Selscan. All 3 were strongly correlated (rho values of 0.99 and 0.55 and p-values of < 0.001 Appendix Figure 6) iHS determined by rehh is shown. The median score was 0.66 (95 % CI 0.03-2.18), 31 SNPs had scores in the top 2.5 % and 13 in the top 1 % of the genome (Figure 6.7). These SNPs highlight 3 loci which are likely recently under positive selection (Table 6.7). The top 2.5 % identified very large loci in the context of the Ct genome, the top 1 % identifies similar but more precise loci including a region covering *tarP* and *pmp* family members and a region within the Ct plasticity zone. Three genes identified under selection by Tajima's D and Fay and Wu's H were within these regions, CT049, CT050 and CT622. There is a clear overlap of regions under selection from these different methods, particularly when comparing Fay and Wu's H and iHS (Figure 6.8).

Table 6.7. Regions under positive selection identified by iHS.

The top 3 regions were identified using the top 2.5 % of SNPs, the second 3 regions were identified using the top 1 % of SNPs.

WINDOW START (KB FROM ORIGIN)	WINDOW END (KB FROM ORIGIN)	REGION LENGTH (KB)	NUMBER OF SNPS	GENES WITHIN REGION
5863	257279	251.42	17	CT005-CT228
258295	536643	278.35	4	CT229-CT456
708681	1015110	306.43	10	CT674-CT937
54660	89333	34.67	2	CT048-CT074
180250	180280	0.03	7	CT154-CT155
536643	708681	172.04	2	CT456-CT625

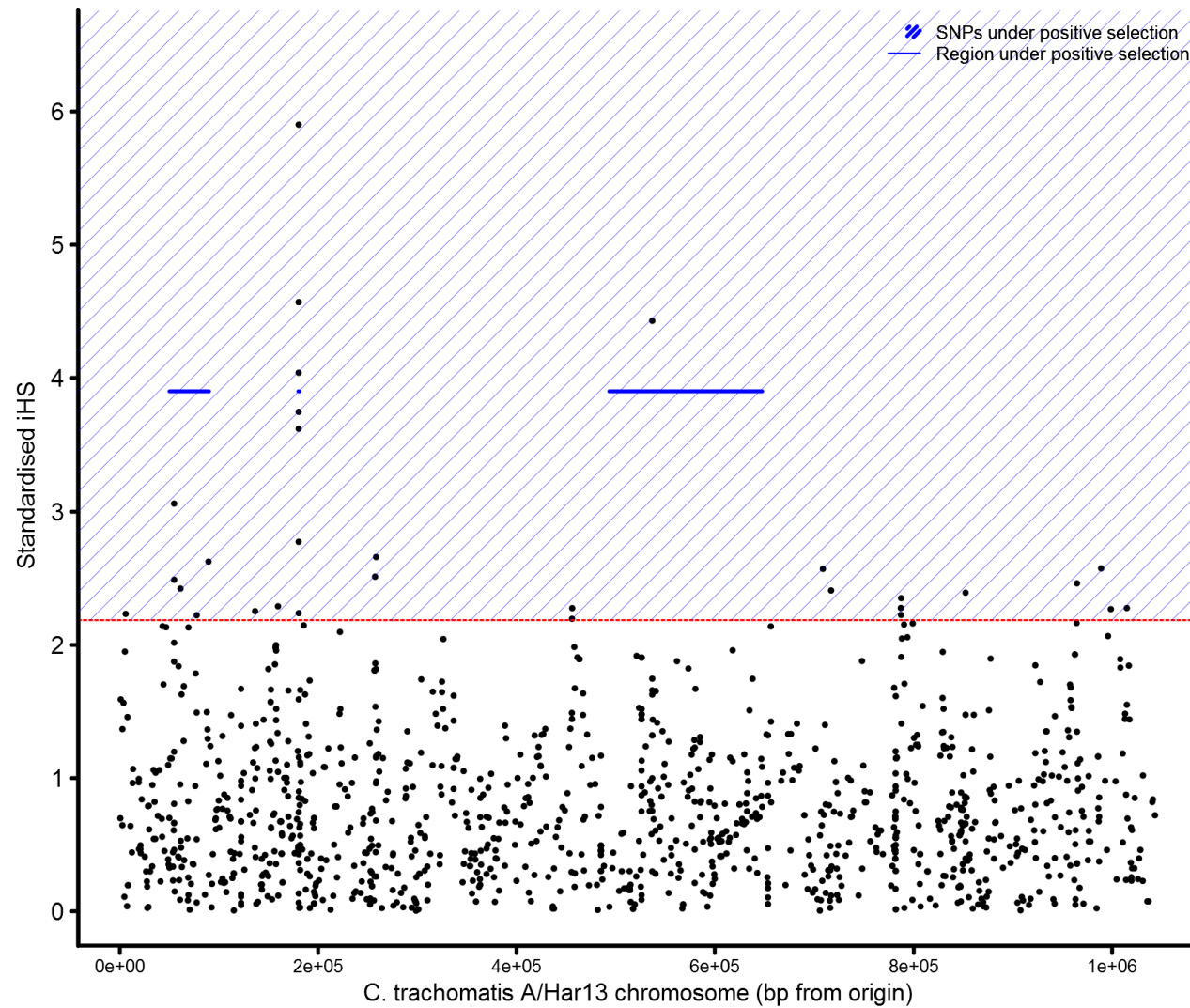


Figure 6.7. Genome-wide evidence of SNPs and loci under positive selection by iHS.

Genes in the top 2.5 % of values are indicated (dashed red line). Regions (blue lines) and genes (shaded blue area) under positive selection are indicated.

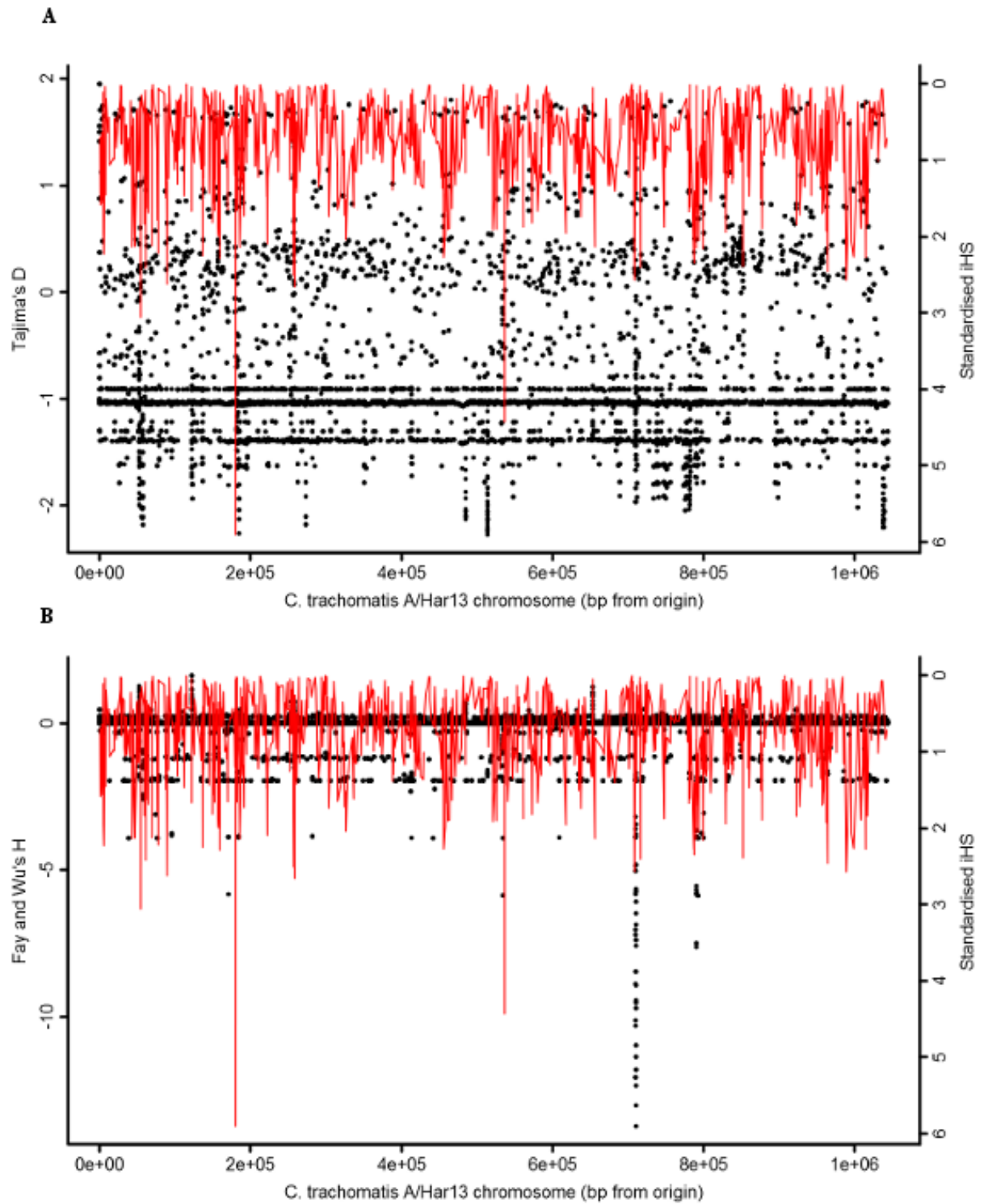


Figure 6.8. Overlap of regions under selection by Tajima's D or Fay and Wu's H with iHS.

There was clear overlap of regions under selection by sliding window analysis of D (A) and H (B) with integrated haplotype scores (red lines).

To support the results, previous data for Ct genes shown to be under selection were used. Due to the lack of Ct populations reflecting isolates currently in circulation, particularly within the trachoma biovar, only one study has previously employed Tajima's D to examine within population diversity. Andreasen *et al*¹⁰⁸ looked exclusively at *ompA* (CT681) in a population of 77 ocular Ct isolates from a trachoma-endemic community in The Gambia. They found significant evidence of selection in *ompA* (D -1.76), these results had a similar but larger deviation from neutrality (D -2.20). A related study of *ompA* genotypes in Tanzania found an overall dN/dS of 0.57 which was indicative of purifying selection¹⁰⁹, this study did not find gene-level evidence of selection however the epitope within the surface-exposed variable domain 1 with the strongest evidence of selection was under purifying selection, D -1.99 and H 0.46. Two further studies tested for selection within the biovars, using just 3 and 9 ocular isolates respectively^{115, 356}. Both studies identified members of the *pmp* family as under positive selection, *pmpB*, *pmpE* and *pmpF* within the trachoma biovar. This study found variable evidence of selection in these genes (D -2.12, 0.08 and -1.16, H -0.21, -3.86 and 0.12). This study did find evidence of purifying and positive selection in *pmpH* and *pmpI* respectively (D -2.85 and -2.57, H 1.99 and -3.47). Borges *et al*³⁵⁶ also identified CT050, CT115 and CT456 under positive selection, these results mostly agreed with this interpretation (D -2.64, -2.33 and -1.62, H -1.66, 0.22 and -6.84). Joseph *et al*¹¹⁵ identified a further 11 genes under selection, all of these had D values less than zero and 6 had significant evidence of selection at either the gene or epitope level in these data.

In this Ct population 48 genes were identified with evidence of selection by either a combination of Tajima's D and Fay and Wu's H, at the gene or epitope level, or by iHS. Secreted and outer membrane proteins were significantly over-represented in these targets, as were genes with peak expression levels very early or very late in the developmental cycle (Figure 6.9). Expression at these pivotal stages of infection and localisations, which would facilitate host interaction support these genes as important factors in Ct survival and pathogenesis.

Fifteen of these genes under selection have been previously identified as immunogenic including; CT681 (*ompA*) the immunodominant major outer membrane protein, CT694 which is utilised in trachoma serology in combination with Pgp3 and CT872 (*pmpH*) another outer membrane protein. This supports our assertion that evidence of selection may help identify genuine host immune targets.

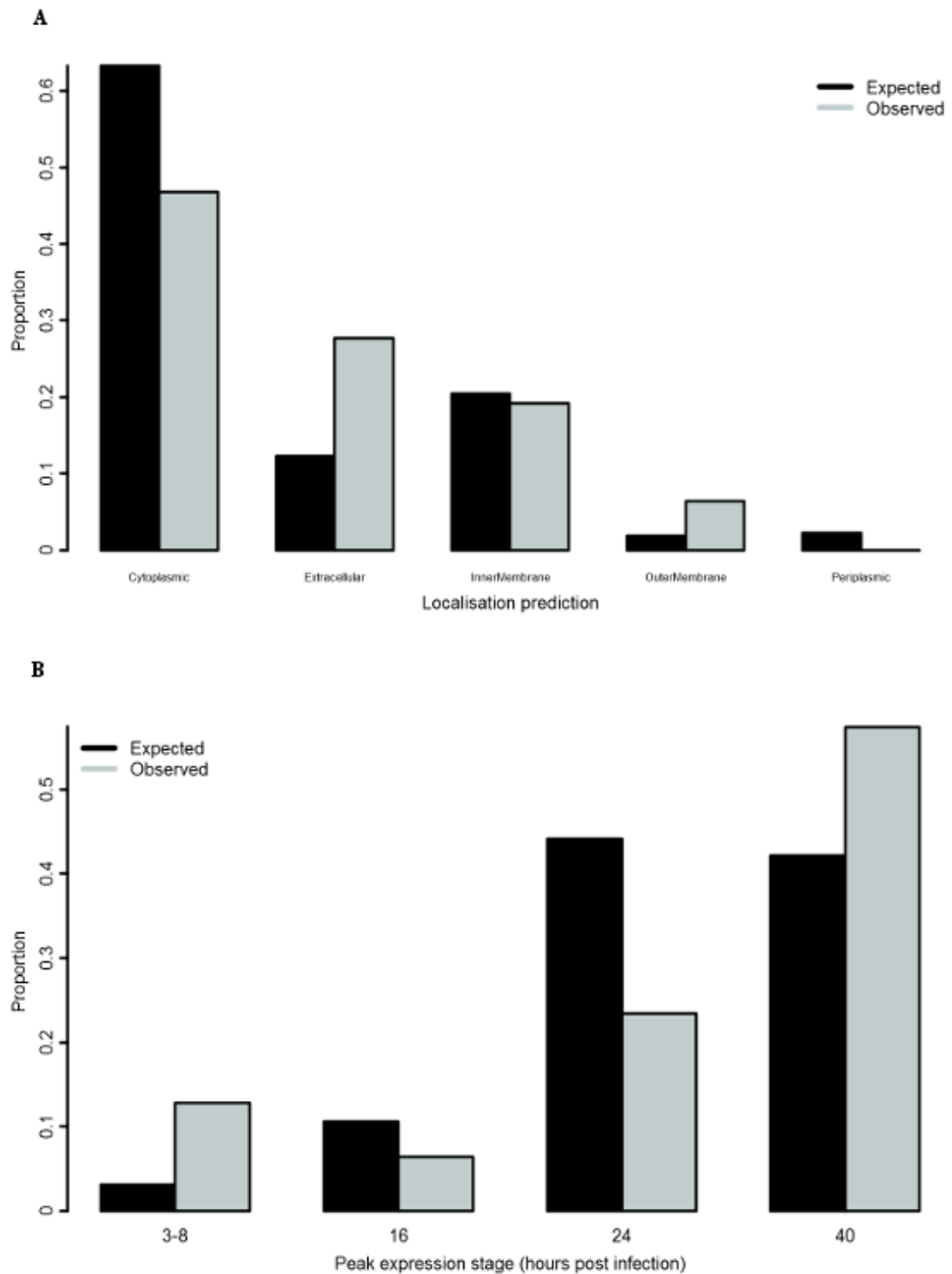


Figure 6.9. Over-representation of genes expressed early or late in the developmental cycle and localised to the outer membrane or secreted in genes under selection.

Proteins identified through transcriptomics as expressed late or very early in the Ct developmental cycle were over-represented in the 48 genes under selection (grey) compared with the whole genome (black). Proteins with a consensus localisation prediction of extracellular or outer membrane were over-represented in the 48 genes under selection (grey) compared with the whole genome (black).

6.2.4. Variable evidence of selection acting on target genes

This Ct population was used to examine for evidence of selection within the list of targets identified in Chapters 4 and 5. The majority of targets from both sections had negative Tajima's D values, only two were above zero but these were not significant (Figures 6.10A and 6.11A). Of those with negative D-values, most had little evidence of selection by Fay and Wu's H (Figures 6.10A and 6.11A). One target, susceptibility-associated CT228, contained a SNP under positive selection by iHS (Figures 6.10B and 6.11B).

Seven of 42 targets associated with susceptibility to infection (chapter 4) had gene-level evidence of selection by Tajima's D of which 4 were supported by Fay and Wu's H, 10 were within regions under positive selection by iHS (Table 6.8). Three were both under selection and within a region under positive selection. CT694 and CT695, had evidence of positive selection at the epitope level supported by D and H values. CT545 and CT806 had evidence of purifying selection at the epitope level. CT570 had sliding-window level evidence of selection by Tajima's D. None of the 5 targets associated with protection from infection had any evidence of selection.

CT314 was the only one of 8 scarring-associated targets which had evidence of selection, positive H values suggested it was likely under purifying selection. CT442, the only target associated with protection from scarring, had strong evidence of selection both at the whole-gene and epitope-level. H values suggested different epitopes within this gene were under positive and purifying selection.

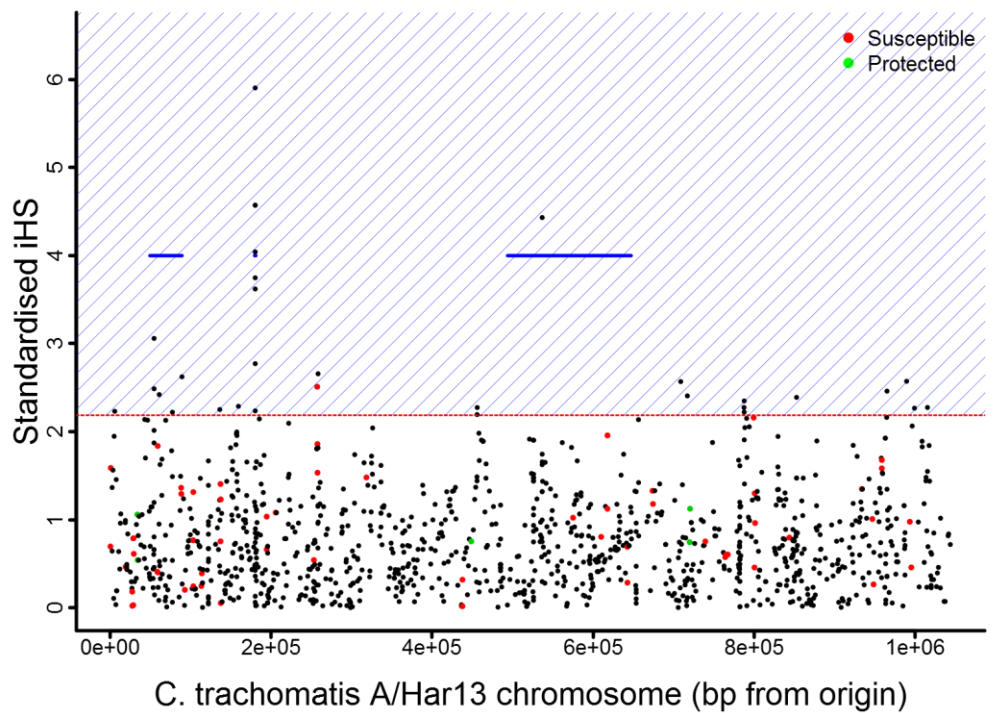
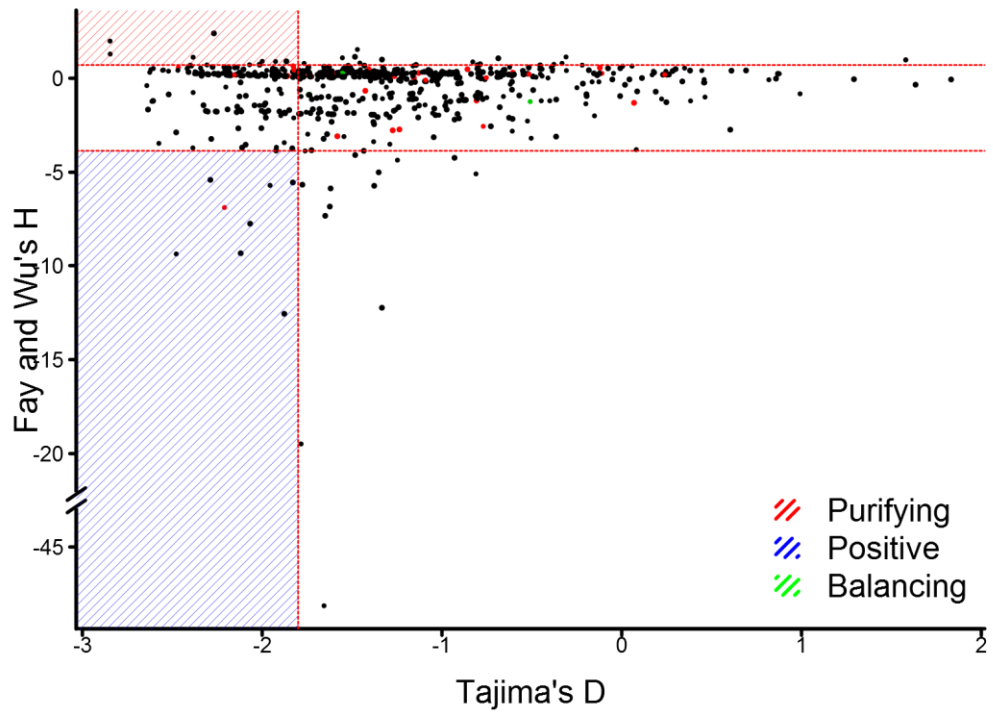


Figure 6.10. Evidence of selection in antibody targets associated with protection from or susceptibility to infection.

Genes under selection by A) D and H values or B) iHS are shown, association with protection from (green) or susceptibility to (red) infection is indicated. Values significantly different from zero are indicated for each measure (dashed red lines). Type of selection is indicated (shaded areas).

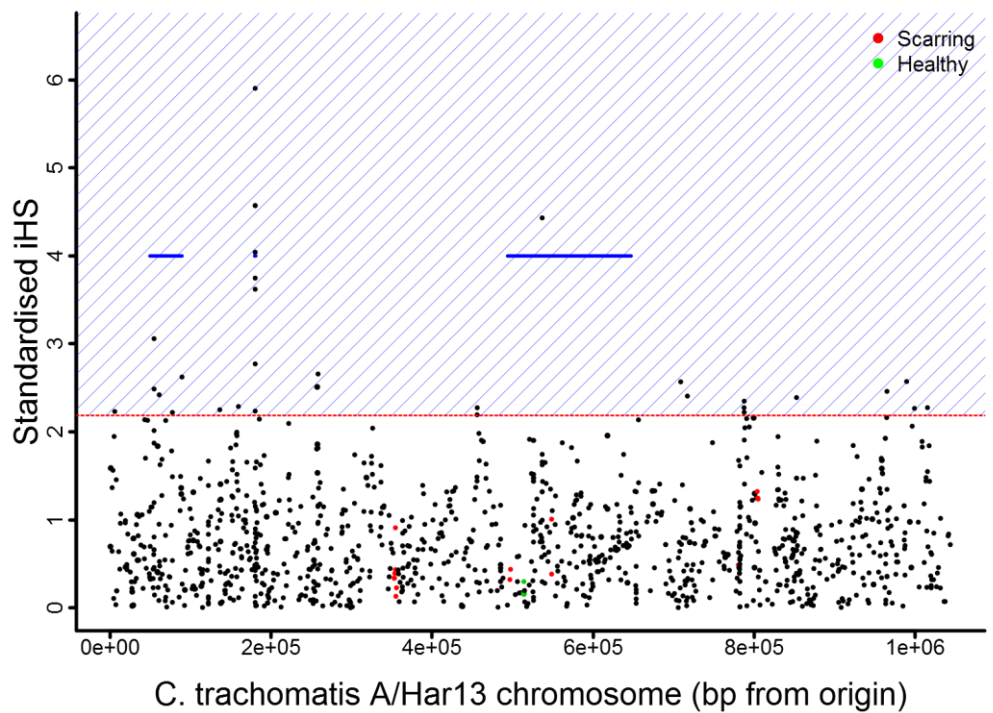
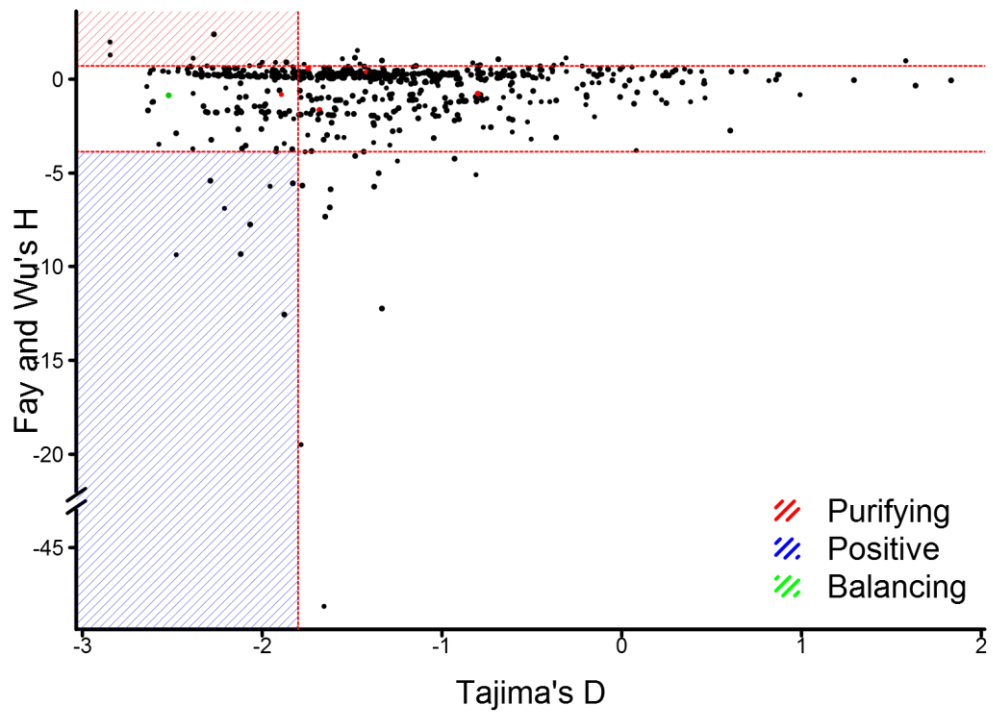


Figure 6.11. Evidence of selection in antibody targets associated the presence or lack of scarring in adults.

Genes under selection by A) D and H values or B) iHS are shown, association with the presence of (red) or lack of (green) scarring is indicated. Values significantly different from zero are indicated for each measure (dashed red lines). Type of selection is indicated (shaded areas).

Table 6.8. Evidence of selection in immune targets identified in chapters 4 and 5.

Genes were ordered by position in the Ct D/UW3 genome. iHS indicates if the gene fell within the three regions identified as under positive selection using the top 1 % of SNPs. Association from chapters 4 and 5 is indicated.

ID	NO. SNPS	THETA	TAJIMA'S D	TAJIMA'S D SW	FAY AND WU'S H	FAY AND WU'S H SW	IHS	ASSOCIATION
CT017	8	0.001	-1.78		0.12		N	Susceptible
CT021	8	0.002	-1.16		0.37		N	Susceptible
CT023	8	0.001	-0.61		0.03	14	N	Susceptible
CT051	11	0.001	-1.43		0.30		Y	Susceptible
CT073	7	0.001	-1.24		-0.66	14	Y	Susceptible
CT078	4	0.001	-1.13		-2.72		N	Susceptible
CT089	7	0.001	-1.27		0.29		N	Susceptible
CT097	5	0.001	-0.86		-2.77	7	N	Susceptible
CT106	5	0.001	-1.74		0.49		N	Susceptible
CT118	4	0.002	-0.12		0.18	19	N	Susceptible
CT119	1	< 0.001	-1.04		0.57		N	Susceptible
CT123	2	0.001	-1.31		0.02		N	Susceptible
CT142	4	0.001	-1.77		0.06		N	Susceptible
CT168	4	0.002	-1.73		0.09		N	Susceptible
CT181	3	0.001	-1.61		0.10		N	Susceptible
CT223	3	0.001	-0.81		0.07		N	Susceptible
CT228	8	0.003	0.069		-1.18		N	Susceptible
CT237	3	0.001	-1.62		-1.30		N	Susceptible
CT284	6	0.001	-1.58		0.07	6	N	Susceptible
CT316	1	< 0.001	-1.03		-3.09		N	Susceptible
CT381	0	0.000	NA		0.02		N	Susceptible
CT494	5	0.001	-1.86		0.00		Y	Susceptible
CT502	1	< 0.001	-1.03		0.13		Y	Susceptible
CT541	2	0.001	-0.11		0.02		Y	Susceptible
CT545	9	< 0.001	-1.82		0.27	14	Y	Susceptible
CT570	6	0.001	-0.75	14	0.44		Y	Susceptible

CT579	8	0.001	-2.15		0.03		Y	Susceptible
CT584	0	0.000	NA		0.18		Y	Susceptible
CT592	11	0.001	-1.83		0.00		Y	Susceptible
CT642	20	0.005	-2.47		0.59		N	Susceptible
CT664	7	0.001	-1.41		0.59		N	Susceptible
CT668	6	0.002	-1.58		0.52		N	Susceptible
CT694	14	0.003	-2.21		0.31	13	N	Susceptible
CT695	7	0.001	-0.77		-6.89	14	N	Susceptible
CT703	3	< 0.001	-1.62		-2.56		N	Susceptible
CT728	2	0.001	-1.38		0.48		N	Susceptible
CT764	2	< 0.001	-1.38		0.42		N	Susceptible
CT795	1	< 0.001	0.24		0.04		N	Susceptible
CT806	12	0.001	-1.83		0.20	5	N	Susceptible
CT813	2	< 0.001	-0.52		0.67		N	Susceptible
CT841	7	0.001	-1.10		0.22		N	Susceptible
CT875	9	0.001	-1.27		-0.09		N	Susceptible
CT029	10	0.003	-1.74		-0.87		N	Protected
CT334	4	< 0.001	-1.79		0.09		N	Protected
CT391	5	< 0.001	-1.55		0.25		N	Protected
CT630	2	< 0.001	-0.51		-1.24		N	Protected
CT314	23	0.001	-1.68		-1.63	14	N	Scarred
CT425	10	0.001	-1.74		0.58		N	Scarred
CT442	28	0.012	-2.52	30	-0.86	17	N	Healthy
CT471	6	0.002	-0.80		-0.76		N	Scarred
CT645	0	0.000	NA		0.00		N	Scarred
CT667	0	0.000	NA		0.00		N	Scarred
CT679	6	0.001	-1.42		0.39	2	N	Scarred
CT698	12	0.002	-1.89		-0.81		N	Scarred
CT706	1	0.000	-1.04		0.02		N	Scarred

6.2.5. Evidence of selection in predicted epitopes

To further investigate the relationship between evidence of selection and host immune targets predicted epitopes and windows under selection were mapped onto the 7 targets with the strongest evidence of selection. Predictions of membranous regions were also included to highlight areas that are less likely to be accessible to the host immune system. *ompA* was used as a positive control to evaluate this methodology as its immunogenic epitopes have been extensively mapped. For *ompA* the positions highlighted in red correspond to the four variable domains (VD)³⁴, known to be surface-exposed and contain the majority of the immunogenic epitopes (Figure 6.12). Three of the 5 regions with non-zero D values were within the VD, however, most were not supported by corresponding H values. The region with strongest evidence of selection was found within VD1, this has a negative D value and a positive H value indicating purifying selection. VD1 is known to be immunogenic and contain a neutralising epitope in the trachoma biovar¹⁰⁷. This suggests host-immune targeting of this region may be driving selection. The only other region with evidence of selection was found toward the C-terminus, this highlighted a transmembrane region which appeared to be under positive selection. Selection pressure in this region is likely not driven by immune targeting.

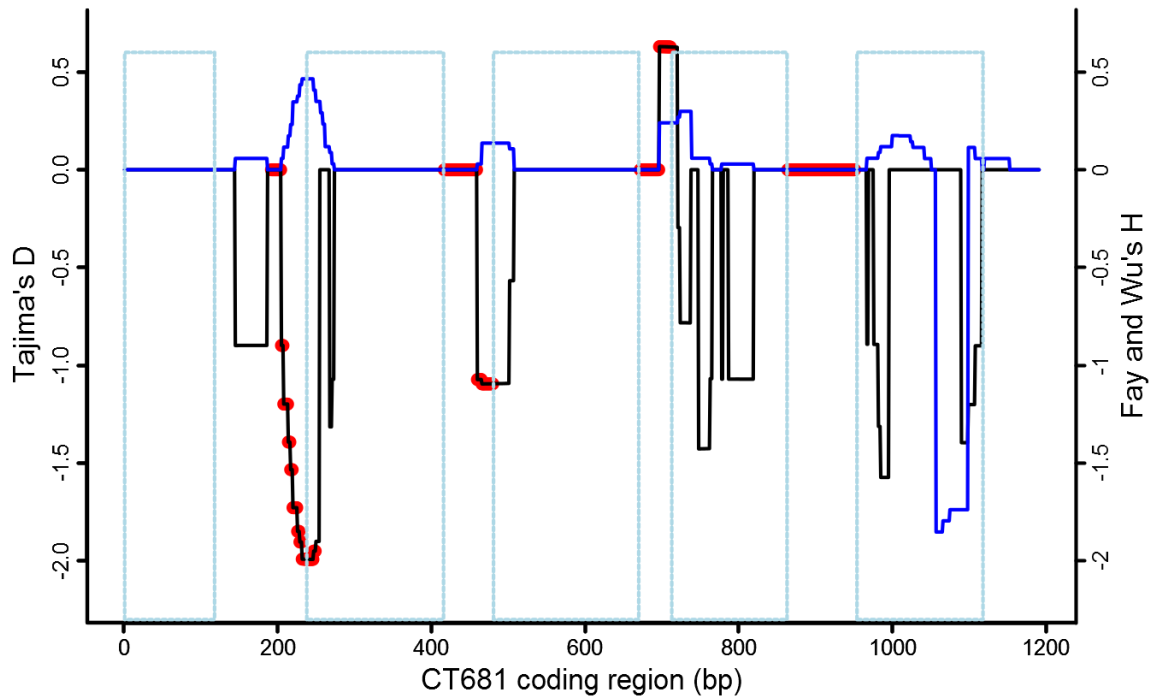


Figure 6.12. Evidence of selection in CT681 (*ompA*).

Values of D (black) and H (blue) across the gene based on sliding windows of 42 nucleotides are shown. Variable domains are indicated (red). Predicted transmembrane domains are indicated (dashed light blue).

This methodology was further validated using CT868 (*Dub1*), a secreted effector identified as under selection in this population. B-cell epitopes were predicted as described previously. A number of regions had evidence of selection by D-values, 2 of these were supported by significantly negative H values (Figure 6.13). The second of these, close to nucleotide 1100, had strong evidence of positive selection and overlapped with 3 predicted epitopes. Similarly to *ompA*, not all selection in this gene was being driven by immune pressure, but combining epitope predictions and evidence of selection identified potentially important immunogenic epitopes around nucleotide 1100.

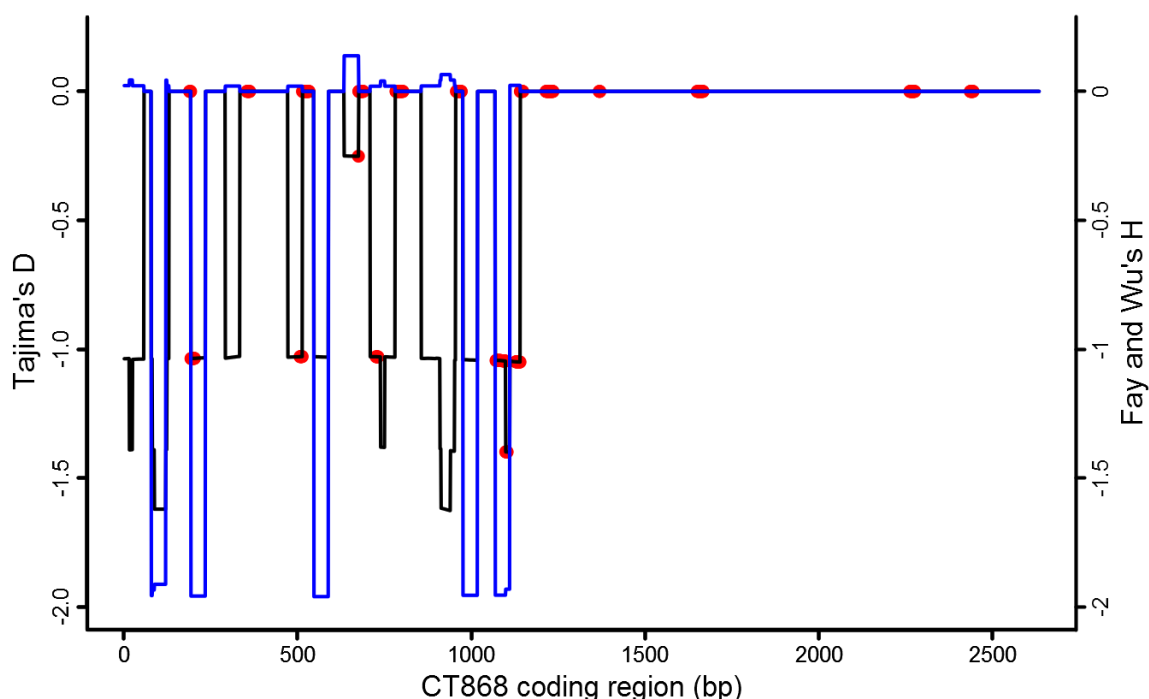


Figure 6.13. Evidence of selection in CT868 (*DubI*).

Values of D (black) and H (blue) across the gene based on sliding windows of 42 nucleotides are shown. Predicted B-cell epitopes are indicated (red).

CT442, which had the strongest evidence of selection in the targets studied in this thesis through expression and ELISA, showed limited overlap between regions under selection and B-cell epitopes (Figure 6.14). The two regions under purifying selection, indicated by a negative D and positive H, contained no predicted epitopes. The region under positive selection also contained no epitope and was predicted to form part of a transmembrane helix. A third region with non-significant negative D value did overlap with a predicted B-cell epitope, this region is around the beginning of the immunogenic epitope used in chapter 7.

CT228 was the only target to have evidence of positive selection by iHS and showed variable overlap between regions under selection and B-cell epitopes (Figure 6.15). Three regions of which one had a significantly negative D value, at the 0.01 level, had no overlap with predicted epitopes. The third region did contain one of the sites identified as under positive selection and had two short predicted epitopes immediately before and after it. The two regions with positive D values approaching significance

both overlapped with predicted epitopes, the second of which had a significantly negative H value suggesting there may be balancing selection acting within this region.

CT694 had extensive regions predicted to be potential B-cell epitopes, these regions had mixed overlap with evidence for selection (Figure 6.16). Three regions were predicted to be under positive selection, two of which were significant by both D and H values. The one non-significant region and one of the significant regions under positive selection overlapped with predicted epitopes, the C-terminal located epitope of these appeared to be under strong selection.

The remaining 4 targets had limited overlap between regions under selection and B-cell epitopes, or they had epitopes predicted across a proportion of their sequence to large be deemed reliable (Appendix Figure 7).

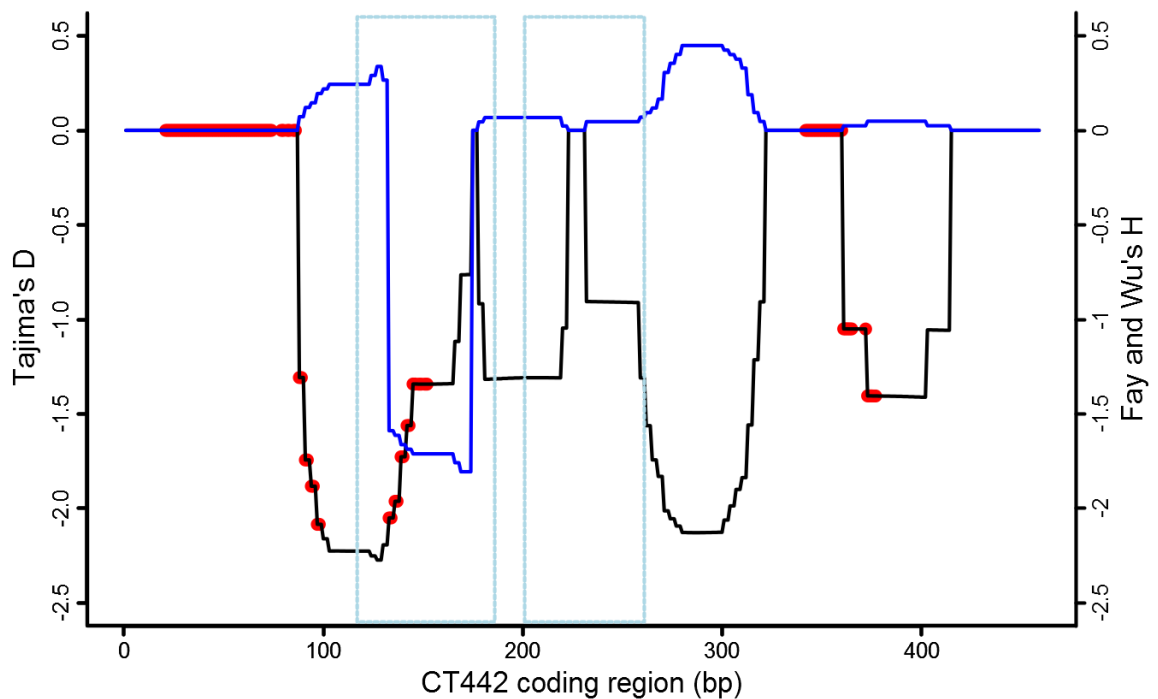


Figure 6.14. Evidence of selection in CT442 (*crpA*).

Values of D (black) and H (blue) across the gene based on sliding windows of 42 nucleotides are shown. Predicted epitopes are indicated (red). Predicted transmembrane domains are indicated (dashed light blue).

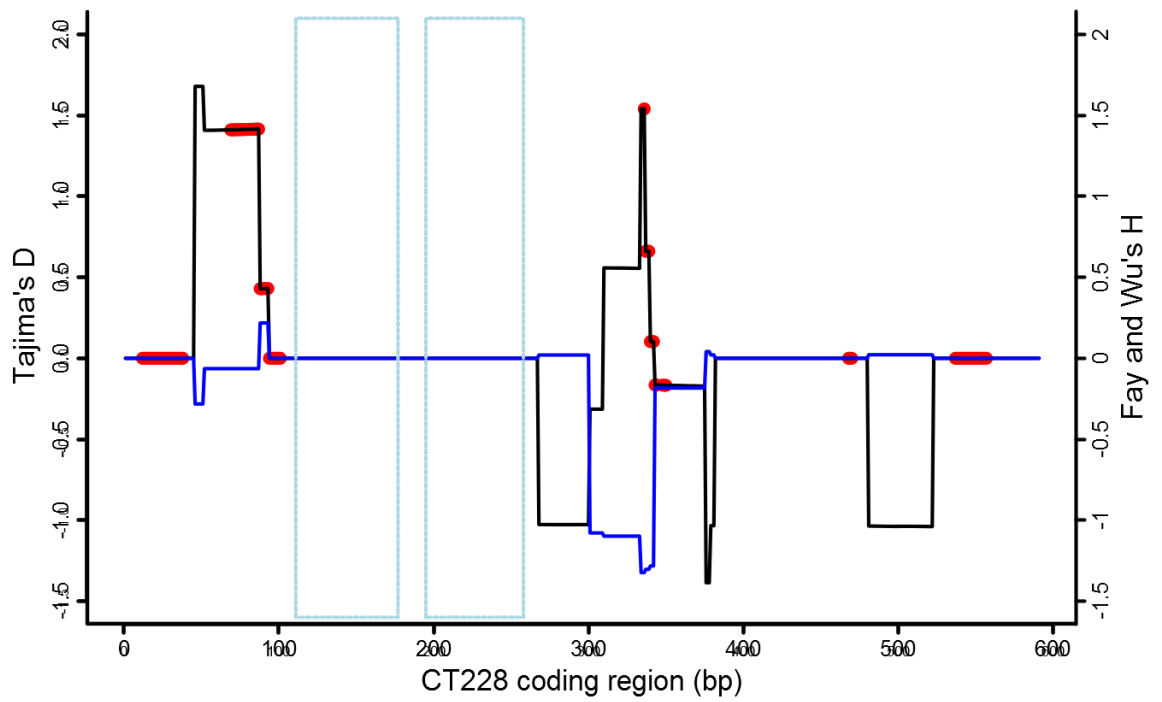


Figure 6.15. Evidence of selection in CT228.

Values of D (black) and H (blue) across the gene based on sliding windows of 42 nucleotides are shown. Predicted epitopes are indicated (red). Predicted transmembrane domains are indicated (dashed light blue).

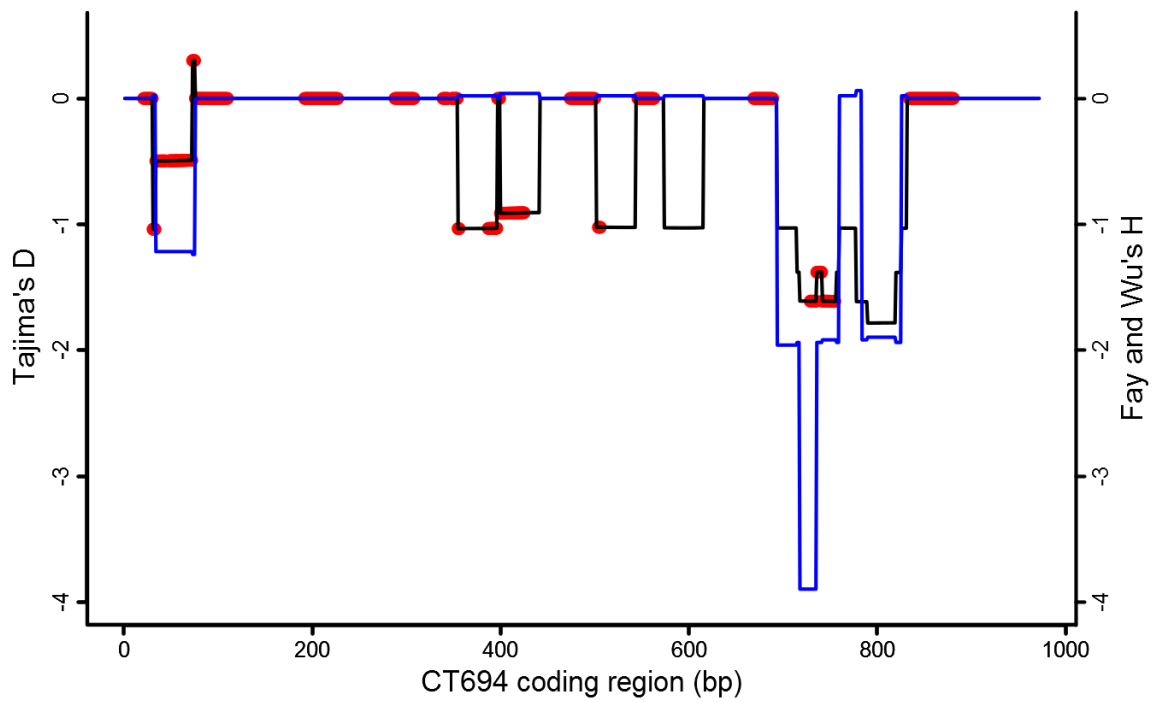


Figure 6.16. Evidence of selection in CT694.

Values of D (black) and H (blue) across the gene based on sliding windows of 42 nucleotides are shown. Predicted epitopes are indicated (red).

6.3. Discussion

This study combined bioinformatics tools for predicting immunogenicity with sequence data from 126 ocular Ct isolates collected from trachoma-endemic communities on the Bijagos Islands, Guinea-Bissau to test for evidence of natural selection and examine potential overlap with known and putative immune targets. Tajima's D, Fay and Wu's H and the more recently described integrated haplotype score were used to test independently for selection and to cross validate results from the different methods. These techniques identified a number of genes and regions of the Ct genome under purifying and positive selection, these regions were focussed in immunogenic proteins and those localised to interact directly with the host. There was limited significant evidence of balancing selection in immunodominant proteins or across the genome, suggesting Ct may employ different strategies for immune evasion than those commonly seen in other pathogenic organisms. Evidence for selection within the potential immune targets identified in chapters 4 and 5 was variable, 7/55 had significant evidence of selection by at least one method. Regions under natural selection overlapped with predicted B-cell epitopes in CT228 and CT694, indicating immune pressure might have been the driving force in variation of these two genes. This overlap was not as clear for the remaining 5 genes, meaning other forces of selection were likely driving the observed genotypes.

6.3.1. Genome-wide evidence of purifying and positive selection

Genome-wide, the median of Tajima's D was -1.39 and the median of Fu and Li's F^* was -2.75. This indicates an excess of low frequency alleles, many of which were found just once in the 126 ocular Ct isolates. The genome-wide median of Fay and Wu's H was 0.06. Combined, this means the majority of genes with negative values of D and F^* had values of H close to zero. This means these low frequency alleles were likely caused by population expansion and are not being driven by natural selection. Therefore, as expected the majority of genes are evolving under neutrality.

In total, 17 genes had significant evidence of selection at the gene-level. Positive selection was acting on 8 of these, indicated by negative values of D and H. Purifying selection was acting on 9 of these, indicated by negative values of D and positive values

of H. A further 6 genes had evidence of positive selection at the epitope-level and 9 had evidence of purifying selection at the epitope level. Only 3 genes had evidence of selection at both levels. CT105 and CT641 were under purifying selection at both levels, CT622 had gene-level evidence of positive selection but contained epitopes evolving under both positive and purifying selection. These results show the importance of searching at different levels for signatures of selection, because opposing signatures of selection in different regions of a gene can mask one another when viewed at the gene-level. Equally a short region under selection could be missed. There was no significant evidence of balancing selection at the gene or epitope-level.

Integrated haplotype scores identified 3 regions of the Ct genome under positive selection based on the top 1 % of SNPs ranked by iHS score; CT048-CT074, CT154-CT155 and CT456-CT625. Peak signals of selection from Tajima's D and more clearly Fay and Wu's H overlapped with high scoring SNPs by iHS. Only 3 genes identified by a combination of Tajima's D and Fay and Wu's H were within these regions; CT049, CT050 and CT622. This discrepancy is likely due to the different comparisons these methods utilise. Tajima's D examines exclusively within the population being studied, comparing the frequency of low and intermediate frequency alleles in genes or windows. Integrated haplotype scores compare haplotypes that have arisen within the population to an ancestral haplotype from outside the population. Fay and Wu's H is related to both measures, comparing the frequency of high and intermediate frequency alleles in genes or windows and utilising an ancestral sequence to determine if alleles have arisen within the population. It is therefore not surprising that Fay and Wu's H correlated with both Tajima's D and iHS, while D and iHS were less strongly correlated.

6.3.2. Secreted proteins, outer and inclusion membrane proteins are under strong forces of selection

Evidence of selection in genes suggests they are important for Ct survival and continued transmission, changes in these genes impacts the genetic fitness of individual isolates providing a selective advantage or disadvantage. It is reassuring that genes that encode outer and inclusion membrane proteins and secreted proteins are over-represented in those under selection in this population. Combined with the over-representation of very

early and late expressed genes, these localisations are pivotal points of interaction with the host. Genes under selection included outer membrane proteins (CT396, CT681 and CT872), type-3 secreted effectors (including CT053, CT456 and CT694), Incs (including CT147, CT223 and CT228) and other known virulence factors (CT868). The plasticity zone, which is known to be variable and important in pathogenicity, also contained three genes under selection. This demonstrates that these host-interacting proteins are important in differences between the biovars but also within biovars and individual populations of Ct.

For genes under positive selection changes are being favoured. These changes could prevent or delay recognition by the immune system or they could improve infectivity or intracellular survival. CT694 is under positive selection in this population. CT694 is a type-3 secreted effector which is thought to be important during Ct invasion of host cells^{63, 362}, it is also a common antibody target in trachoma-endemic populations²⁸⁶. Mutations in CT694 may delay immune recognition of isolates with epitopes a host hasn't been exposed to previously, providing a selective advantage. Equally for this gene and CT456 (*tarp*) which is also under positive selection, mutations could alter their binding efficiency for their host targets. These proteins are both important in early invasion events³⁶³, as such small changes could impact the initial events in Ct intracellular survival providing mutated isolates an advantage.

For genes under purifying selection, polymorphism is being driven out of the population. These changes could be in conserved genes where an essential function is impacted. CT223 is under purifying selection in this population. CT223 is an Inc involved in Ct reprogramming of the host microtubule⁸⁷, without which bacterial infectivity is severely reduced. Mutations in this gene may impact its ability to recruit the host protein CEP170 which is required for Ct control of microtubules. Purifying selection is not normally seen in immunodominant antigens, but if responses against these antigens benefitted the bacteria then it is possible that mutations changing the sequence would be selected against. As described previously, antibodies against MOMP and LPS can block neutralising antibodies directed against PmpD *in vitro*¹³⁴. If this phenomenon occurs *in vivo* it is plausible that Ct isolates with common MOMP or LPS

epitopes that hosts frequently recognise and generate blocking antibodies against would be at a selective advantage. Variable domain (VD) 1 of *ompA*, the gene which codes for MOMP, had evidence of purifying selection in this population. This domain has been shown to be immunogenic³⁶⁴. Therefore, antibodies targeted against this region may be providing a selective advantage in this population, which could be related to the *in vitro* demonstration of blocking antibodies.

Anti-MOMP antibodies can also enhance Ct infectivity through FcγRIII *in vitro* and in mice^{133, 365, 366}. Ctad1 is involved in EB attachment and induction of host-cell signals required for invasion²⁹⁹. It is plausible that antibodies binding Ctad1 could target EBs to host cells or directly induce uptake through cross-linking of Fc receptors.

6.3.3. Limited evidence of balancing selection and Ct immune evasion

Balancing selection is a common method of immune evasion employed by human pathogens^{343, 350, 367, 368}. This form of selection maintains multiple allelic forms of immunodominant genes in a population such that when immune responses recognise and clear isolates with one allele it provides a selective advantage for isolates presenting a different allele. This process continues cyclically as each allele becomes common and is therefore selected against, with the simultaneous selective advantage for all other alleles maintaining diversity however there was no significant evidence of balancing selection in this population. The genes identified as potentially under balancing selection based on Tajima's D after lowering the threshold for significance were not supported by Fay and Wu's H. Ct is a successful pathogen, able to repeatedly reinfect individuals within households and communities, immunity develops slowly and may only be partial, therefore Ct must utilise different strategies for immune evasion.

One possibility is that antibodies against some EB surface antigens can block protective and neutralising antibodies as discussed previously. In this scenario antigens which induce blocking antibodies should be under purifying selection. Antigens which induce neutralising antibodies should be under positive selection, because even if they

are being partially blocked the development of neutralising antibodies against an isolate would put them at a disadvantage. In *ompA* there is evidence of purifying selection in a known immunogenic region, but *pmpD* has no significant evidence of positive selection.

Another possibility is the decoy hypothesis described in chapter 4, where Ct presents an extensive panel of irrelevant antigens to the host immune responses. Antibody responses against these are not protective and actively divert responses away from potentially protective epitopes. This would possibly drive purifying selection in some of the non-protective antigens due to the advantage they provide in evading clearance by the host immune response, however since it proposes that a wide array of antigens are involved there may be not strong signs of selection pressure in any one gene.

6.3.4. Evidence of selection in susceptibility and scarring associated antigens identifies candidate immune targets

Of the 55 potential immune targets identified in chapters 4 and 5, only 7 had evidence of selection in this Ct population. This supports the ‘decoy’ hypothesis because 37/42 of the antigens associated with susceptibility to Ct infection had no evidence of selection and therefore are not individually inducing responses which provide a selective advantage or disadvantage. Two of the susceptibility associated antigens which had evidence of selection were under purifying selection, which is also possible in this hypothesis. The three susceptibility associated antigens with evidence of selection were CT228, CT694 and CT695. Regions under positive selection in CT695 did not overlap with predicted B-cell epitopes, as a secreted effector, functional interactions with the host may be driving this selection.

CT694 is an immunodominant antigen in trachoma-endemic populations, but also functions early in Ct invasion of host cells. It is thought to be involved in reversing some of the actin cytoskeleton changes induced by TARP upon entry. CT694 had two regions with evidence of positive selection, one of which overlapped with predicted epitopes. This suggests that this gene may be under selection driven by immune responses and differences in Ct intracellular survival and transmission.

CT228 is an Inc involved in regulating release of CT from host cells⁹⁹. CT228 had evidence of positive selection by iHS and weaker evidence of balancing selection by Tajima's D and Fay and Wu's H. The region potentially under balancing selection overlapped with a predicted epitope, whereas the region under positive selection did not contain a predicted epitope. Similarly to CT694, these signatures of selection may have separate driving factors. The most likely way for CT228 as an Inc to be targeted by antibodies would be through exposure on the surface of extrusions¹⁰¹. It is possible that the function of CT228 in directing Ct cell exit to either lysis or extrusion could be related to the development of antibody responses against itself. Extrusion provides a selective advantage for CT because it prolongs extracellular survival and protects EBs from immune recognition. If CT228 is exposed on the surface of extrusions antibodies targeting this protein could enhance CT clearance and therefore select against extrusion or at least isolates which contain a CT228 allele commonly recognised by the host immune responses. A combination of these factors may be driving the varied selection seen in CT228.

CT314 is found in the outer membrane complex of Ct EBs³²⁹, antibody responses were associated with TS in adults. CT314 was under purifying selection in this study. This would fit with either function postulated in chapter 5. Antibodies targeting this antigen could enhance infectivity via Fc receptor-mediated uptake or they could block neutralising activity of protective antibodies. In both situations responses against CT314 would provide a selective advantage to isolates presenting a commonly recognised epitope, therefore driving out diversity through purifying selection. There was no observed overlap of regions under selection and predicted B-cell epitopes, therefore immune responses may not be the driving force for selection in CT314.

CT442 is an Inc, antibody responses were higher in children whose conjunctival scarring had not progressed in the previous 4 years. One region within CT442 was under positive selection in this population, however this was found within a predicted transmembrane domain and therefore this is unlikely to be immune driven. Two regions were under purifying selection, this does not fit with the protective effects of CT442 antibodies and these regions did not overlap with predicted B-cell epitopes. These strong signatures of selection in CT442 do not seem to be immune response-driven. The functional importance of CT442 is not known, the results from this Ct population suggest mutations in its sequence may have negative effects on Ct survival and transmission.

6.4. Conclusions and future work

Signatures of selection in this population of Ct isolates from the Bijagos Islands, Guinea-Bissau were concentrated in genes with critical functions in cell entry, intracellular survival and continued transmission. There was evidence of positive and purifying selection, but limited balancing selection was detected. Ct evasion of host immune responses therefore does not rely on balancing selection in a few immunodominant antigens. Evasion may instead utilise previously described blocking antibodies, which interfere with potentially neutralising antibodies, or a mechanism described here as the decoy hypothesis, where antibody responses against a large panel of non-protective antigens diverts the immune response away from potentially protective epitopes. Seven of the potential immune targets identified in chapters 4 and 5 had evidence of selection, although most of these are also functionally important for Ct survival therefore selection is not definitively immune response-driven.

Future work should focus on validation of genes identified as under selection in independent populations of Ct isolates. Genes under selection driven by impact on Ct survival should be common to ocular Ct populations, those under immune response-driven selection may vary depending on the levels of transmission and endemicity. The apparent absence of balancing selection in this population and its implications for Ct evasion of the host immune responses should also be investigated further. Both through whole-genome approaches as described here but also through targeting sequences of immunodominant antigens believed to be important in immunity to Ct, as has been done previously for *ompA*.

7. Functional studies of CT442 a candidate antigen associated with protection from scarring trachoma

7.1. Introduction

7.1.1. Functional importance of inclusion membrane proteins

Inclusion membrane proteins (Incs) are a family of type-3 secreted effectors unique to *Chlamydia*, despite having minimal sequence homology they have a common bi-lobed hydrophobic region approximately 60 amino acids in length which is believed to span the inclusion membrane⁸³. Fifty-five to 62 Ct proteins are predicted to be Incs, of which 23 are shared across the genera^{301, 369}. Approximately half of these have been demonstrated to localise to the inclusion membrane, several have been localised within the inclusion and some remain undefined³⁷⁰. Incs account for 2.7 % of the Ct genome and 4 % of its coding capacity, suggesting they are important for Ct survival.

Host organelles including Golgi-derived vesicles⁸⁴, multi-vesicular bodies⁸⁵, lipid droplets⁸⁶, rough endoplasmic reticulum⁷⁶ and centrosomes⁸⁷ are recruited to the inclusion and are important in its development, acquisition of nutrients and Ct survival and transmission. Until recently only a few Incs had been functionally investigated, reviewed in chapter 1. In 2015 Mirrashidi *et al* performed large-scale affinity purification-mass spectrometry on 58 predicted Incs, host interactions were identified for 38 and showed global functionality in endocytosis, ubiquitination, apoptosis, cell division and DNA damage/repair⁹⁶. This methodology did not identify all known Inc interactors but it did provide a broad map highlighting organelles and pathways previously identified as important in Ct intracellular development (Inc-interactome). CT442 was one of the Incs with defined host interactors from this study.

7.1.2. Limited previous characterisation of CT442

CT442 was first identified as a 15 kDa band from polyacrylamide gels of serovar L2 Ct-infected HeLa cells³⁷¹. It was originally defined as a cysteine-rich protein (crpA) of the outer membrane despite not being cross-linked by disulphide bonds as seen with known Ct outer membrane proteins³⁷². Subsequent studies showed it only contained four cysteine residues¹⁹⁵, the crpA nomenclature has frequently remained in use in the literature. Expression was confirmed in urogenital and ocular Ct isolates, serovar E and

C strains were shown to have 17 and 14 single nucleotide polymorphisms (SNPs) compared with an L3 strain^{373, 374}. These SNPs meant antibodies against different regions of CT442 were both biovar and species-specific *in vitro*³⁷². In early studies, CT442 was first detectable at 30 hours post-infection (HPI) and continued to increase up to 40 HPI³⁷¹. More recent transcriptomic studies defined it as mid-late expressed, it was detectable as early as 8 HPI with peak levels at 18 HPI and steady expression beyond this point^{212, 213}. This pattern of expression was consistent between strains from each biovar but overall levels dropped significantly from LGV to urogenital and from urogenital to ocular isolates³⁷⁵. CT442 expression was strongly down-regulated during IFN γ -induced persistence at all time-points post-infection¹⁰⁵.

CT442 was predicted to be an Inc based on the classical bi-lobed hydrophobic secondary structure pattern and this has been independently confirmed in cell-culture models^{83, 370, 376}, although currently no publication has shown localisation data. CT442 is unique to *Chlamydia* and its function is unknown³⁷⁷. Recent data from the Inc-interactome⁹⁶ and identification of CT442 in association with lipid-droplets during Ct infection⁹³ have provided some clues towards its possible function(s). In the latter, lipid droplets recruited and eventually translocated into the inclusion were examined by mass-spectrometry, this showed an upregulation of lipid biosynthesis in Ct infection and potential interactions with a subset of Inc-s. Depletion of lipid droplets significantly reduced Ct progeny formation. Despite identification of few strong interactors with CT442, enrichment-analyses from the Inc-interactome showed an upregulation of host targets including kinases and proteins involved in vesicular and phagosomal pathways.

7.1.3. CT442 is a cellular and humoral immune target under natural selection

CT442 has been studied relatively extensively with regards to Ct immunology and its recognition by the host immune system. A *Chlamydia* expression library was screened to identify peptides that stimulated Ct-specific CD8⁺ T-cells from murine Ct infections, after sequencing of reactive peptide pools and retesting, a peptide from CT442 (AA 63-72) was identified which stimulated lytic activity from Ct-specific CD8⁺ T-cells¹⁹⁵. These CT442-specific T-cells accounted for 4 % of splenic CD8⁺ T-cells at their peak 6/7 days post-infection, increasing frequency of these cells was negatively correlated with Ct infection forming units (IFU) per spleen. Immunisation with a Vaccinia virus

vector expressing this peptide induced high levels of CT442-specific memory T-cells and reduced IFU per spleen equivalently to immunisation with Ct. Reactivation of CT442-specific memory T-cells was observed upon secondary Ct challenge, despite Ct-induced dampening of memory T-cell responses³⁷⁸.

Generation of antibodies was demonstrated *in vitro* as mentioned above³⁷², *in vivo* anti-CT442 antibodies were first demonstrated from a study of 99 women with current urogenital Ct infection¹⁶⁹. CT442 responses were detectable in over half these women. This finding was corroborated by screening of serum from women with and without tubal factor infertility against a micro-array of Ct antigens(TFI)¹⁶⁸. Responses were higher in women with TFI, however the Ct status and history of the patient groups was unknown therefore these results may have simply reflected previous exposure. C-terminal derived CT442 peptides were also shown to be immunogenic and variably able to distinguish between *Chlamydia* species²⁰⁴ (personal communications, Dr Bernhard Kaltenboeck). In a non-human primate model of trachoma animals that were partially immune to Ct re-challenge showed a marked antibody recall response, this included recognition of a 15 kDa protein from Ct¹³⁶. This was not proven to be CT442 but previous evidence suggests it is likely. In chapter 5 it was found that antibody responses against CT442 were often heightened in individuals protected from scarring trachoma and scarring progression in two geographically and culturally distinct trachoma-endemic populations.

7.1.4. Polymorphism and selection in CT442

CT442 has greater polymorphism than the majority of Ct genes and most Incs^{369, 374}. In chapter 6 it was discovered that CT442 had the second highest per site nucleotide diversity from a genome-wide scan of 126 ocular Ct isolates. Three regions of the gene were identified as under different forms of natural selection. Two regions under purifying selection were located, one within the cytosolic C-terminus and one overlapping the cytosolic N-terminus and first predicted transmembrane domain. A single region under positive selection was found entirely within the first predicted transmembrane region. These showed limited overlap with predicted B-cell epitopes, suggesting the driving force for selection in these regions may not be recognition by antibodies. The region under positive selection does partially overlap with the defined

CD8⁺ T-cell epitope, suggesting host cellular immunity may be driving selection. Six amino acids in CT442 were previously shown to under positive selection exclusively within invasive lymphogranuloma venereum (LGV) isolates³⁷⁵. One of these falls within the region under positive selection, while another lies within the C-terminal region under negative selection. This supports previous evidence from small numbers of Ct isolates that different ecological niches of Ct lead to distinct signatures of natural selection³⁵⁶. This highlights the value of within-population genomic studies to understand variation in Ct survival and pathogenicity.

The two publicly available serovar C Ct genomes both have truncating mutations in CT442³⁷⁹. C/TW3, originally isolated from a Taiwanese conjunctival sample in 1959, has a single adenosine inserted after nucleotide 39. The resulting frameshift introduces a stop codon at amino acid 22. C/UW-1, originally isolated from a conjunctival sample in Seattle, Washington in 1965, has a four nucleotide deletion after nucleotide 39. The resulting frameshift also introduces a stop codon at amino acid 23. More detailed analysis of C/TW3 identified a further fifteen pseudogenes of which 9 have been seen before in ocular strains. Serovar C strains have rarely been found in contemporary trachoma endemic communities or in sub-Saharan Africa trachoma endemic communities^{107-109, 380, 381}. They also appear to possess no detectable *in vitro* growth defects^{382, 383}. It is plausible that truncated genes and as yet unidentified SNPs are factors in the relatively low detection of serovar C ocular Ct isolates in current trachoma endemic populations.

Previous studies and work in chapters 5 and 6 have shown that CT442 is targeted by both cellular and humoral immune responses in ocular and urogenital Ct infections. In LGV isolates and a trachoma-endemic population of Ct isolates, variation in CT442 appears to be driven by natural selection which is only partly due to host immune pressure. As an Inc, CT442 is expressed and localised to the inclusion membrane which facilitates interactions with the host during the intracellular stage of the Ct developmental cycle. It is plausible that some of the signatures of selection evident in CT442 are driven by its as yet undefined role in Ct intracellular survival and transmission. In this chapter cell-culture models of Ct infection were used investigate CT442 localisation to the inclusion membrane and identify host proteins and pathways with which it interacts.

7.2. Results

7.2.1. CT442 sequence homology and predicted structure

CT442 is predicted to have a bi-lobed double transmembrane structure, typical of the majority of Incs. The structure consists of 40 cytosolic amino acids at the N-terminus followed by the first 23 amino acid transmembrane helix passing through the inclusion membrane, after a short intra-inclusion linker region the second 22 amino acid helix passes back through the membrane with a further 60 cytosolic amino acids at the C-terminus (Figure 7.1).

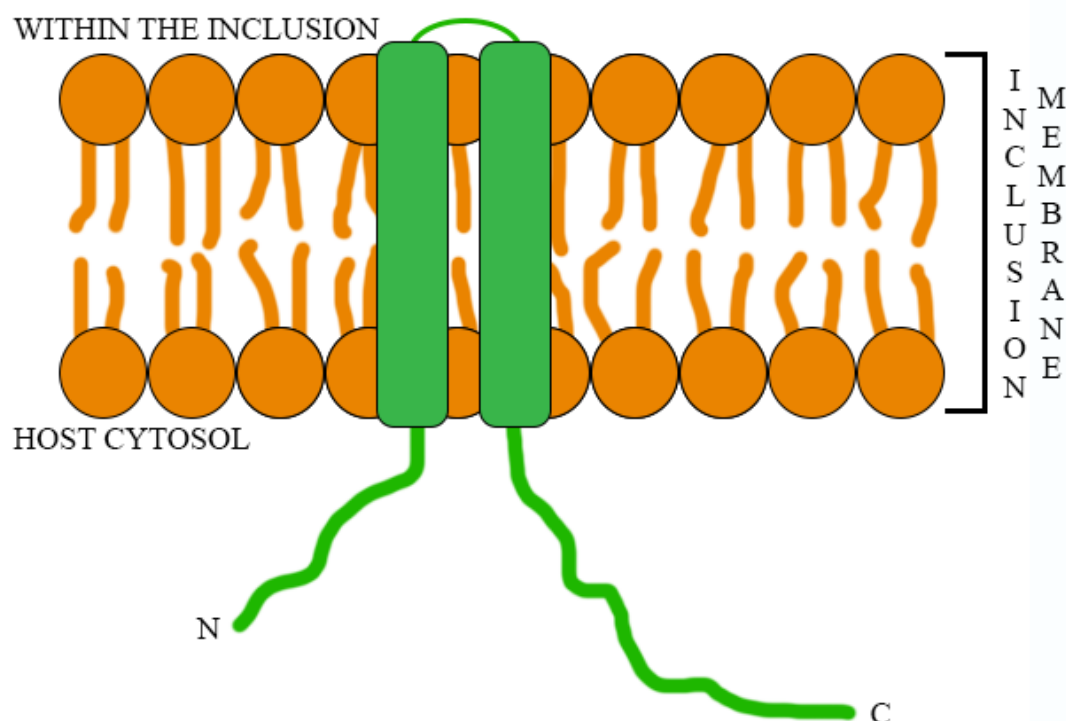


Figure 7.1. Schematic of the predicted structure of CT442.

This representation of CT442 was based on predicted transmembrane domains.

Homologues of CT442 are found within all strains of Ct with polymorphisms found in both membranous and cytosolic regions (Figure 7.2), the mutations causing expression of a truncated protein in serovar C (TW3 and UW-1) are indicated by the gap in the nucleotide alignment just prior to nucleotide 50 (Appendix Figure 8). A truncating mutation was uncovered in one of the Ct isolates collected in Guinea-Bissau,

9471_4_86. Unlike serovar C this was caused by a single nucleotide polymorphism at nucleotide 300 that truncated CT442 after amino acid 99. This isolate could potentially still express both transmembrane domains but would not have the majority of the C-terminal cytosolic domain.

There is variable CT442 (crpA) homology in other *Chlamydia* species; 64 % *C. suis*, 58 % *C. muridarum* and 44-45 % in *C. abortus*, *C. caviae*, *C. felis* and *C. psittaci*. *C. muridarum* and *C. suis* orthologues of CT442 are homologous across the complete sequence. The remaining species have homology from amino acids 29 to 125, covering the predicted transmembrane regions and parts of the cytosolic domains.

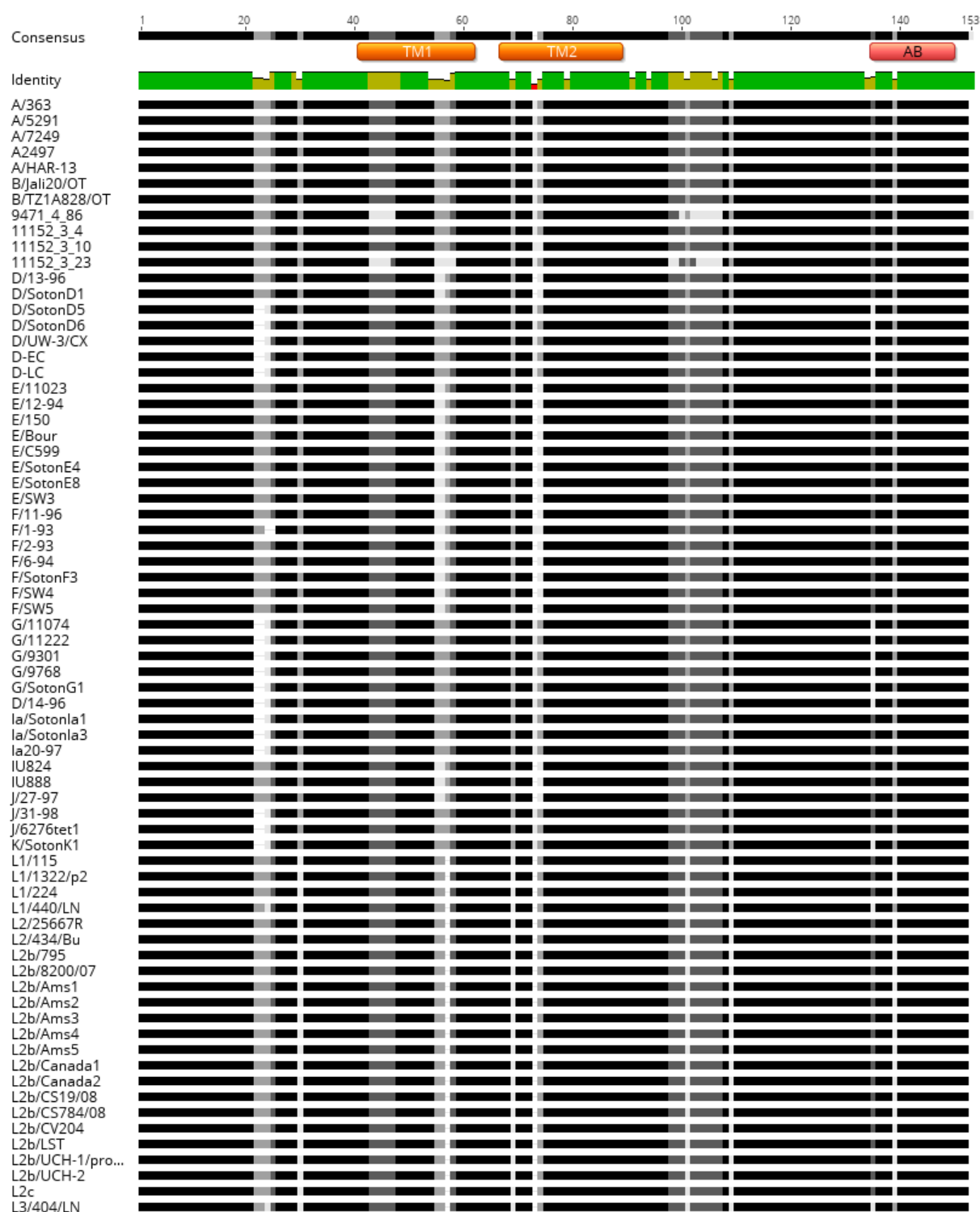


Figure 7.2. Amino acid alignment of CT442 from serovar-representative strains.

The two predicted transmembrane domains are indicated (TM1 and TM2). The region against which the anti-CT442 antibody was generated is indicated (AB). The ‘identity’ sequence shows relative conservation of sequence, from identical (green) to increasing levels of variation (lighter shades of green). This was produced using Geneious.

7.2.2. CT442-His construct design and expression

In addition to the CT442 GST-fusion utilised for immunological work, there was a need for different CT442 constructs. CT442-green-fluorescent protein (GFP) for localisation studies (already been produced) and a CT442-C-terminal six-histidine (His) tag for further immunological studies hopefully resembling more native conformation,

Production of a CT442-His construct was problematic in both cloning and subsequent expression of soluble protein. The problems in cloning were not an unusual feature for chlamydial membrane proteins, particularly those expressed in the inclusion membrane which have caused similar problems. To overcome these difficulties a CT442-His construct was produced by GenScript (Piscataway, NJ, USA).

Colony PCR of the CT442-His construct transformed into competent *E. coli* produced bands of appropriate size, pET22b had no bands (Figure 7.3A). Expression of the CT442-His protein was confirmed by comparing IPTG-induced and non-induced cultures of inoculated-LB, a band just below 17 kDa was close to the expected size of 16 kDa (Figure 7.3B). Growth in LB produced better expression than TB or YT medium, although much of the protein was aggregated after bacterial cell lysis and centrifugation (Figure 7.3C). All detectable protein was aggregated from initial growth and induction at 37 °C for 3 hours with 1 mM IPTG (Figure 7.3D).

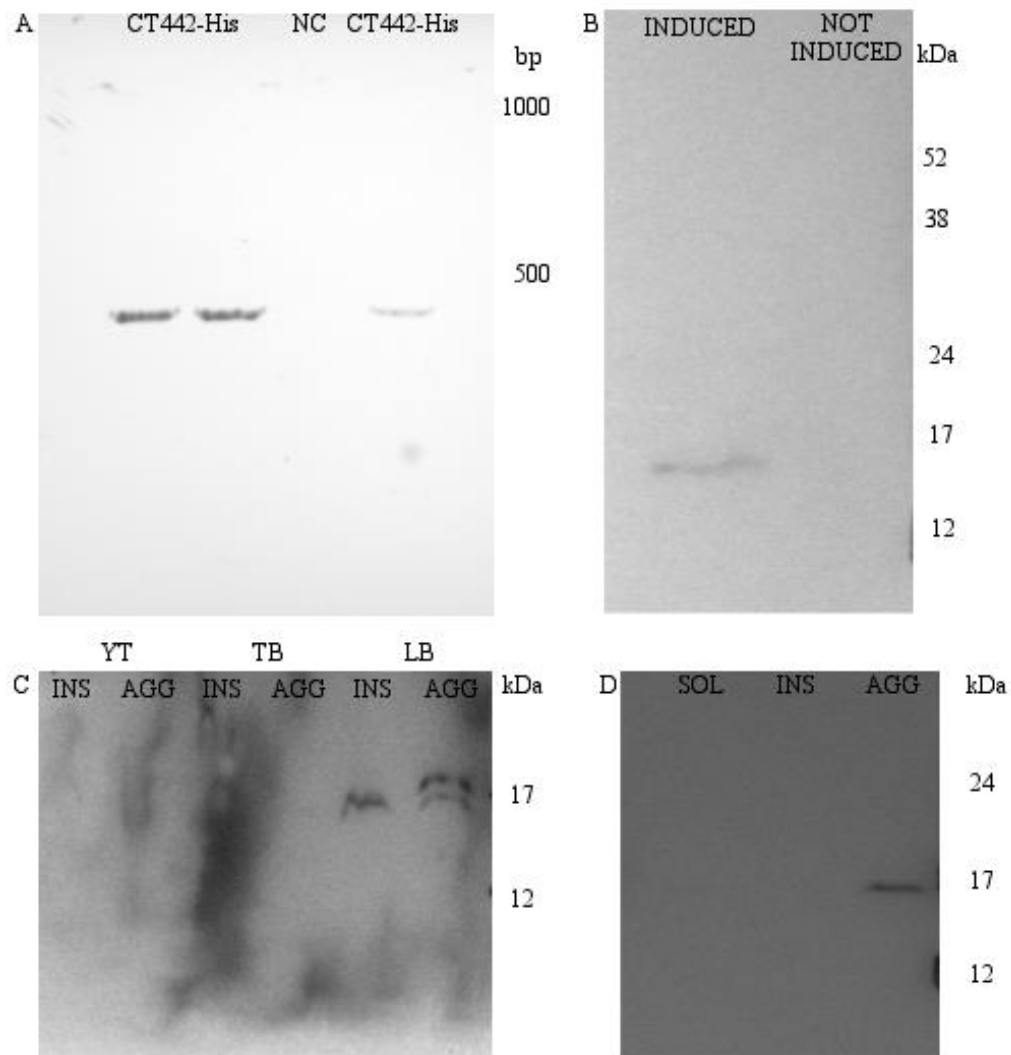


Figure 7.3. Expression trials of CT442-His.

Presence of the CT442-insert in the construct was confirmed by colony PCR (A) from colonies on LB plates (lanes 1 and 2) and a frozen stock of transformed *E. coli* (lane 4). The pET22b⁺ plasmid without the CT442 insert was included as a negative control (NC). Western blots were incubated with an anti-His monoclonal antibody to bind the six-histidine tag. The lysates were separated by ultra-centrifugation into aggregate (AGG), soluble (SOL) and insoluble fractions (INS). B) Bacterial lysates from cultures grown and induced at 37 °C for 3 hours with 1 mM IPTG or no IPTG showed successful expression of CT442. C) Expression of CT442 was best when grown with LB media, protein was aggregated and insoluble in a culture volume of 50 ml. D) CT442 was entirely aggregated grown in a culture volume of 1 L.

Protein expression was tested at 30, 20 and 10 °C for 2, 4 and 16 hours with 0.1, 0.2, 0.6 and 1 mM IPTG. No CT442-His was detectable when induced at 10 or 20 °C. CT442-His was detectable at 30 °C. After 2 and 4 hours regardless of IPTG concentration CT442-His was in the aggregate and insoluble fractions as expected for a membranous protein (Figures 7.4A and 7.4B). After 16 hours CT442-His was concentrated in the aggregate fraction, expression was highest with 0.2 mM IPTG (Figure 7.4C).

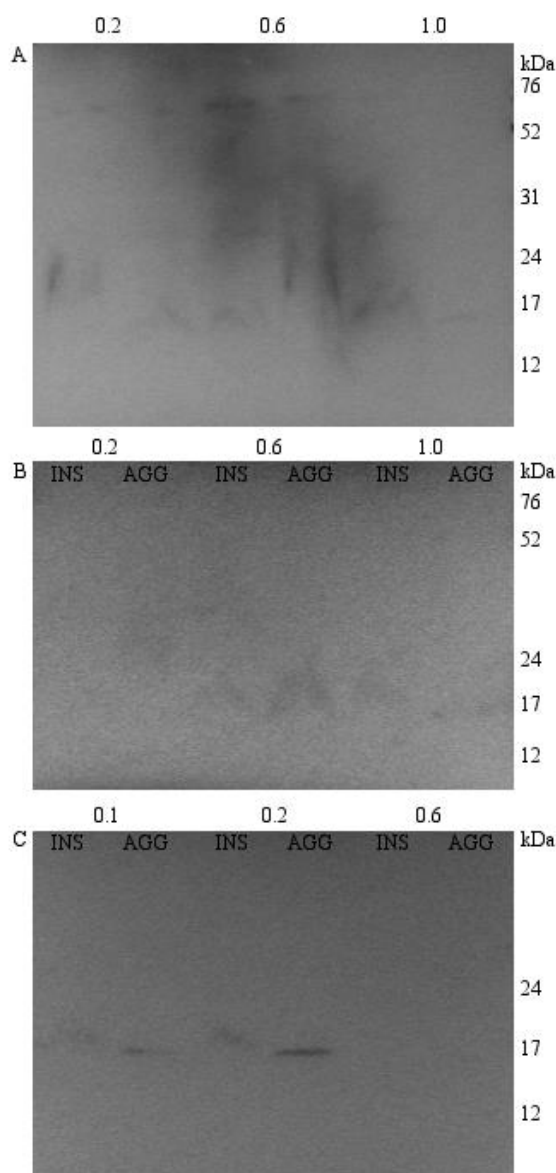


Figure 7.4. Induction trials of CT442-His.

Western blots were incubated with an anti-His monoclonal antibody to bind the six-histidine tag. The lysates were separated by ultra-centrifugation into aggregate (AGG), soluble (SOL) and insoluble fractions (INS). Protein expression was tested at 30 °C for 2 (A), 4 (B) and 16 (C) hours. IPTG concentration in mM is indicated above each blot.

Further conditions were trialled at 30 °C for 4 hours, IPTG concentration was reduced to limit protein aggregation. The protein was primarily found in the aggregate and soluble fractions with both 0.1 and 0.6 mM IPTG, aggregation was significantly reduced with 0.1 mM (Figure 7.5A). This unexpected result was confirmed by repeating the experiment with 0.1 mM, however, with 0.6 mM expression was undetectable (Figure 7.5B).

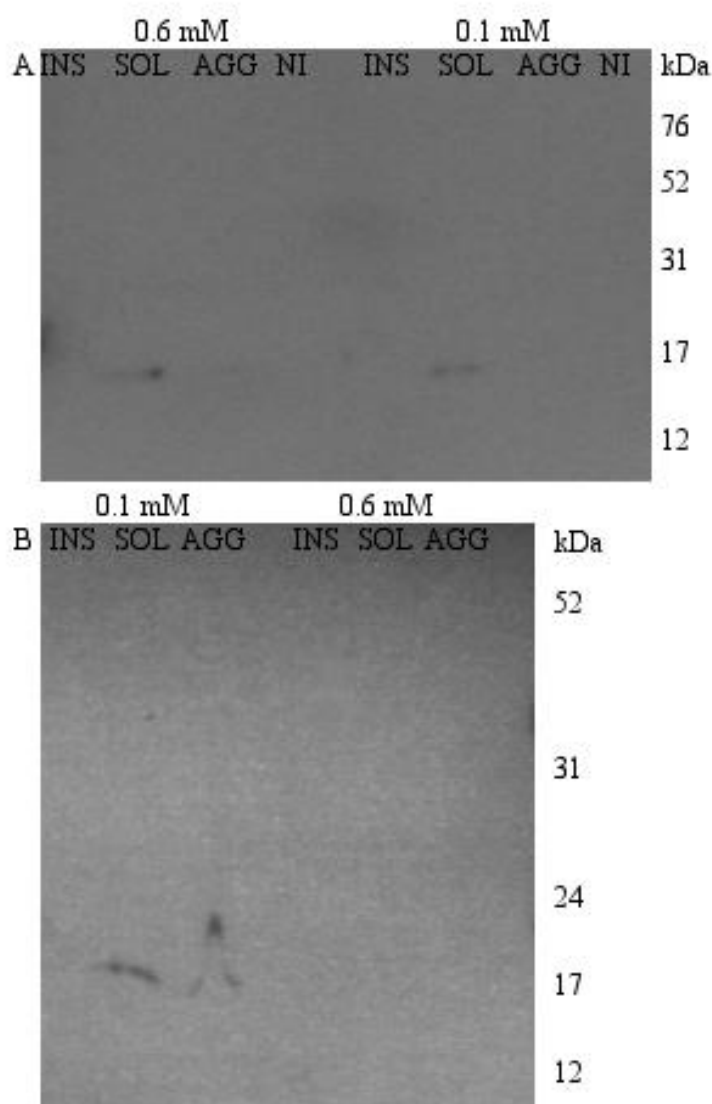


Figure 7.5. Further induction trials of CT442-His.

Western blots were incubated with an anti-His monoclonal antibody to bind the six-histidine tag. The lysates were separated by ultra-centrifugation into aggregate (AGG), soluble (SOL) and insoluble fractions (INS). Non-induced cultures were included as controls (NI). IPTG concentration in mM is indicated above each blot. Protein expression was tested at 30 °C for 4 hours and CT442 was detected in the soluble fraction (A). This soluble expression was repeatable (B).

Purification of CT442-His from the soluble fraction was attempted by affinity chromatography using a HisTrap and ÄKTA purifier size exclusion chromatography system. Four elutions from 3 peaks on the chromatograms from 2 runs were examined (Appendix Figure 9 parts A and B). Bands around the expected size for a CT442-His monomer and dimer respectively were detectable, however they were very faint and were possibly present in pre-elution column washes (Figure 7.6A). The expected band was not detectable by Coomassie staining, many bands were detectable by silver staining but a dominant band at the expected size was absent (Figure 7.6B).

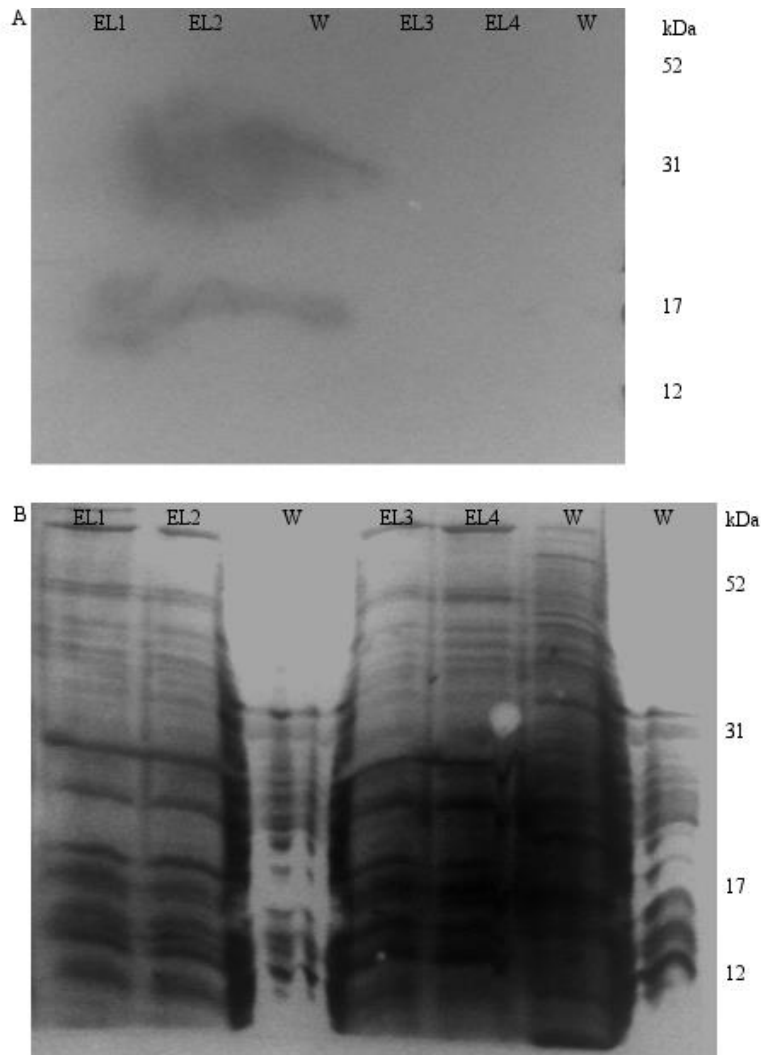


Figure 7.6. Size exclusion-based chromatography for CT442-His purification.

Western blots were incubated with an anti-His monoclonal antibody to bind the six-histidine tag. A) No clear bands were detectable in the elution from peaks in absorbance (EL) or column washes (W). Some unclear bands were observed close to the expected molecular weight, they were present in elution and washes. B) Silver staining of the same samples showed no clear CT442 band in elution, but numerous non-specific impurities.

To improve expression of the protein plasmids designed to express a number of molecular chaperones were tested. These chaperone-plasmids, specifically pKJE7 and pTf16, were transformed into competent *E. coli* and used to make stocks of transformed-competent cells. The chaperone-expressing plasmids were transformed into *E. coli* and these competent cells were transformed with the CT442-His construct.

Expression was not detectable with the pKJE7 plasmid, expression with the pTf16 plasmid was similar to the construct alone, mostly resulting in aggregate with some recombinant protein appearing soluble (Figure 7.7A). HisTrap purification from this soluble fraction produced 2 clear peaks on the chromatogram (Appendix Figure 9C), however no protein was detectable in these 2 elutions (Figure 7.7B). Following this, attempts to express CT442-His were halted due to time-constraints and the need to focus on other areas of investigation.

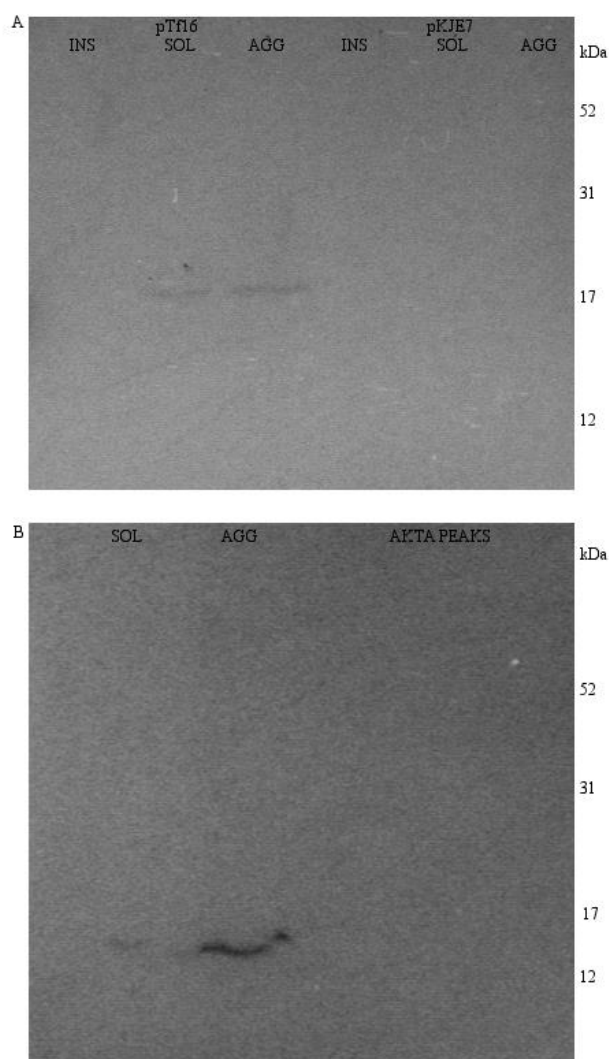


Figure 7.7. Co-expression of CT442-His with molecular chaperones.

Western blots were incubated with an anti-His monoclonal antibody to bind the six-histidine tag. The lysates were separated by ultra-centrifugation into aggregate (AGG), soluble (SOL) and insoluble fractions (INS). A) CT442 was detectable and partially soluble with pTf16. B) No CT442 was detectable in peaks of absorbance from size exclusion-based chromatography.

7.2.3. CT442-GFP localised to the endoplasmic reticulum

Published experimental data and secondary structure homology to known Incs strongly suggested that CT442 is an Inc, however there are as yet no images or formal experimental proof. For purposes of independent verification, the first step in examining the biology of CT442 was the determination of its intracellular localisation. To ectopically express CT442-GFP in HeLa cells, the CT442-GFP construct was transfected into the cells using TurboFect (Thermo Scientific).

CT442-GFP transfected into uninfected HeLa cells localised primarily in the endoplasmic reticulum (ER) membrane, demonstrating its membrane localisation as seen with known Incs (Figure 7.8). Occasionally CT442 formed vesicular structures in the cytosol, which are believed to be related to the acquisition of lipids. In Ct infected HeLa cells the protein also localised to the ER membrane, but the ER focussed at the inclusion membrane (Figure 7.9). This was similar to the results seen with another inclusion membrane protein, IncB, and in contrast to NUE and GFP which maintained their staining of the nucleus and whole cell respectively (Figure 7.8 and Appendix Figures 10 and 11).

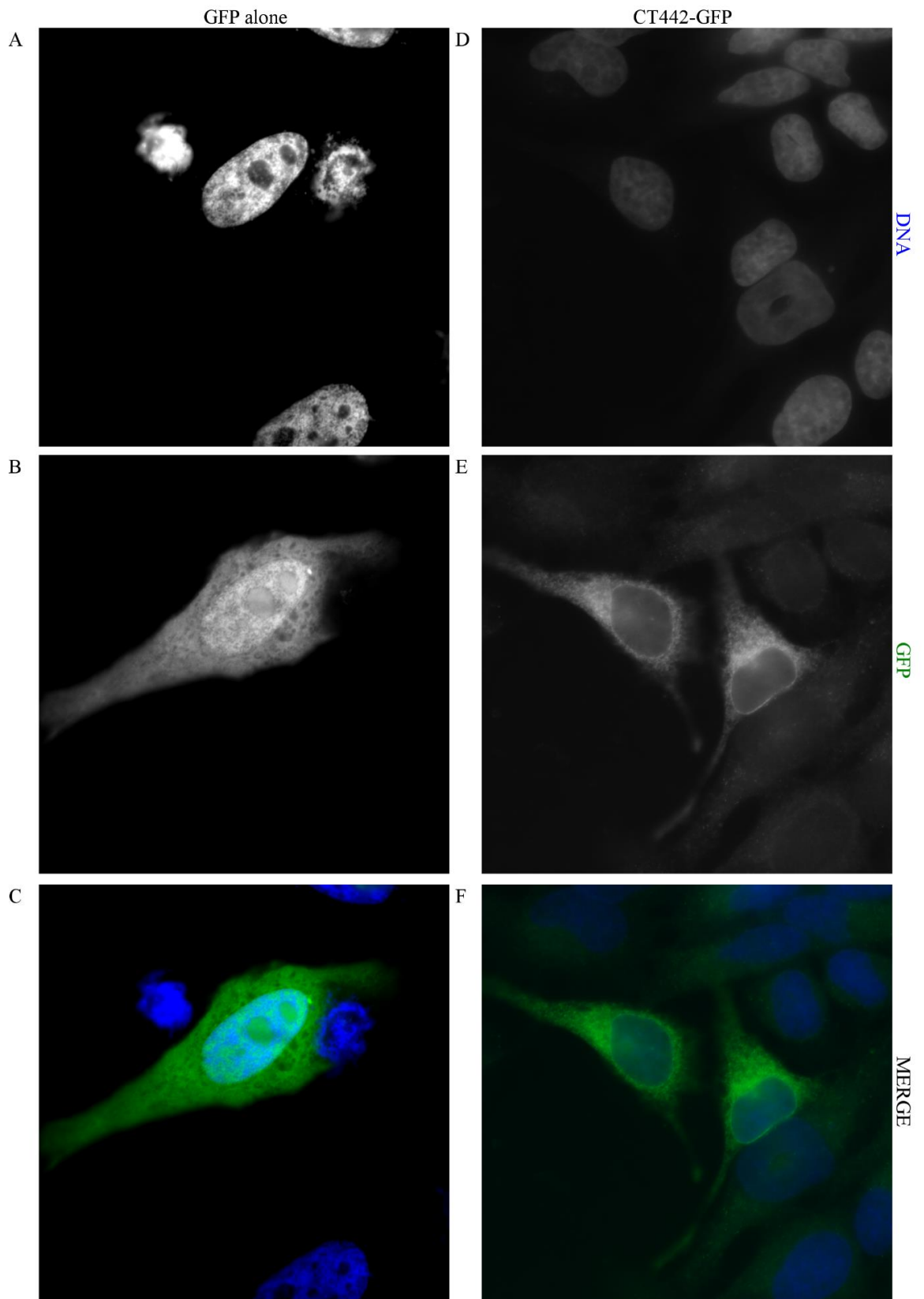


Figure 7.8. Localisation of GFP and CT442-GFP in HeLa cells.

HeLa cells were transfected with GFP and CT442-GFP and fixed after 24 hours. Cells were stained for DNA (A, blue in merged panel) and GFP (B, green in merged panel).

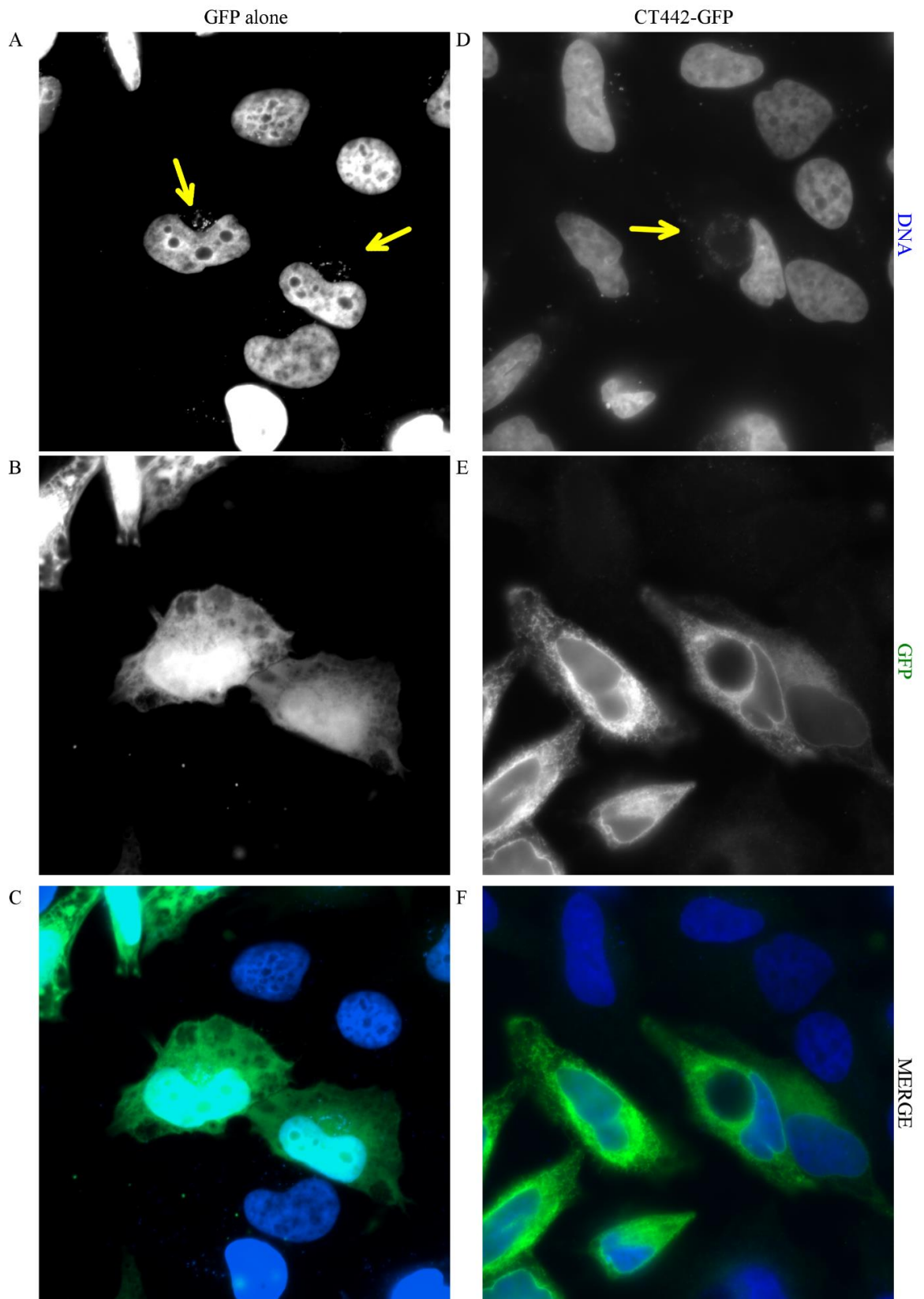


Figure 7.9. Localisation of GFP and CT442-GFP in Ct-L2 infected HeLa cells.

HeLa cells were transfected with GFP and CT442-GFP and infected with Ct-L2, cells were fixed 24 hours post infection (HPI). Cells were stained for DNA (A, blue in merged panel) and GFP (B, green in merged panel). Inclusions are indicated (yellow arrows).

ER-membrane localisation was confirmed by co-staining with an ER marker (calreticulin). CT442-GFP expression in transfected cells co-stains with the ER-marker in uninfected and Ct-infected cells (Figure 7.10).

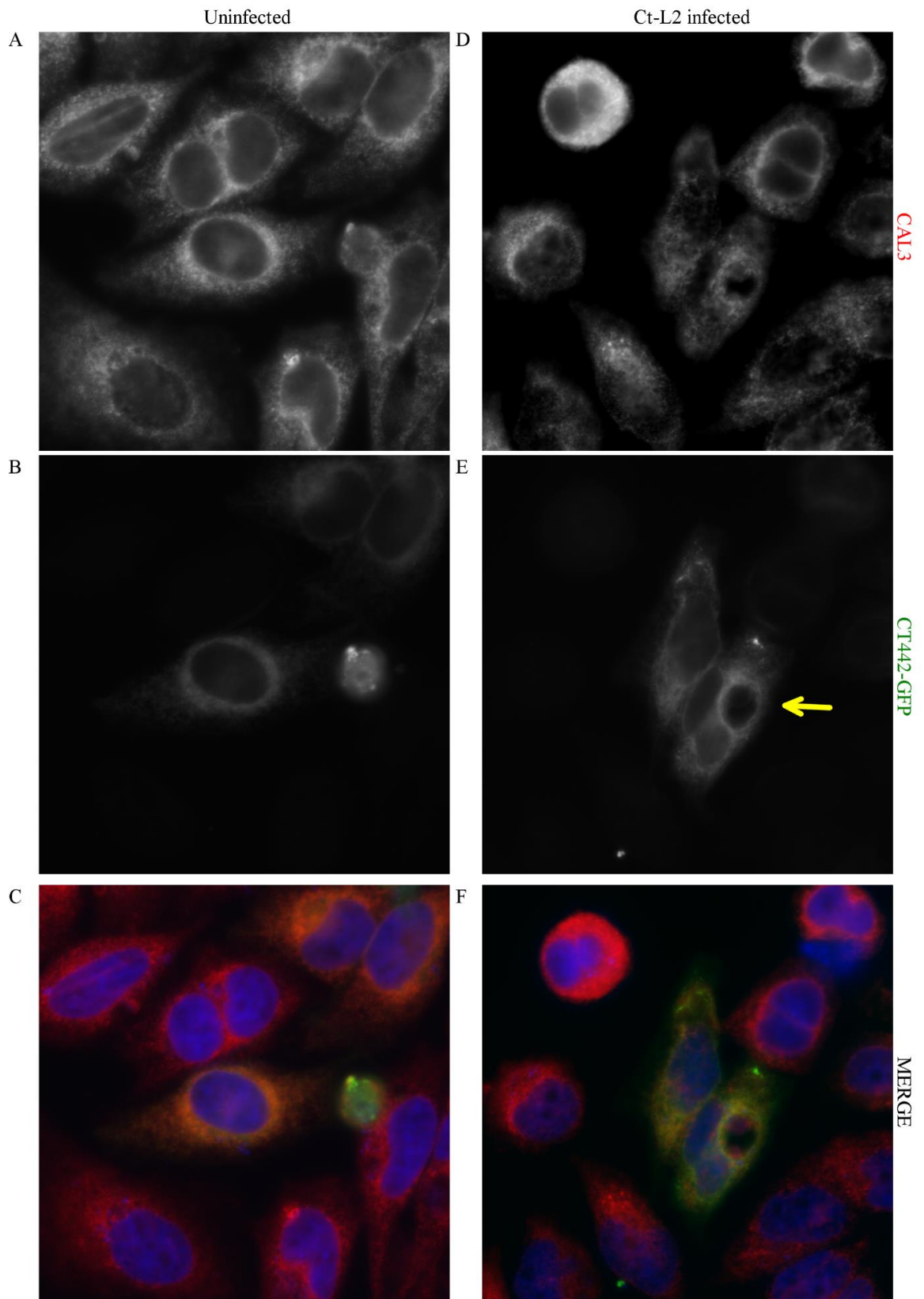


Figure 7.10. Localisation of calreticulin and CT442-GFP in HeLa cells.

HeLa cells were transfected with CT442-GFP and fixed after 24 hours or then infected with Ct-L2 and fixed 24 HPI. Cells were stained for DNA (blue in merged panel), calreticulin (A, red in merged panel) and GFP (B, green in merged panel). Inclusions are indicated (yellow arrows).

7.2.4. CT442 localised in the inclusion membrane from mid-cycle onwards

The CT442 peptide synthesised for serological work in chapter 5 was selected based on previously demonstrated immunogenicity of this region (personal communications, Dr Bernhard Kaltenboeck), this peptide was used to produce CT442 antisera to examine endogenous localisation of the protein. For this rabbits were immunised with the peptide and anti-CT442 antibodies were affinity purified (Genscript). Pre-immune serum was collected as a control. Specificity of the antibody to CT442 was confirmed with lysates from HeLa cells transfected with CT442-GFP (Figure 7.11).

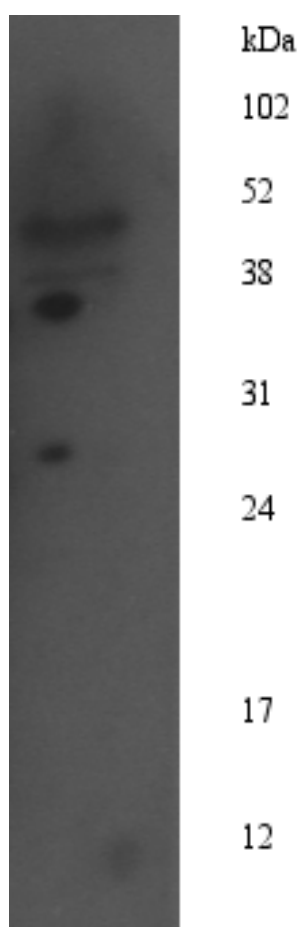


Figure 7.11. Anti-CT442 antibody binds CT442 from Ct-infected HeLa cells.

Western blots were incubated with anti-CT442 antibody to bind CT442. Lysates from CT442-GFP transfected HeLa cells were used to confirm the specificity of the antibody for CT442. The dominant band was between 38 and 52 kDa, the predicted molecular weight of CT442-GFP is 42 kDa.

Initially Ct-LGV2 infected HeLa cells fixed in 4 % PFA were stained with the anti-CT442 antibody at a dilution of 1/100, there was clear staining of the inclusion membrane but also significant background staining (Figure 7.12A). Addition of a BSA blocking step before primary antibody staining significantly reduced background staining (Figure 7.12B), further dilution of anti-CT442 to 1/200 reduced non-specific staining without diminishing staining of the inclusion membrane (Figure 7.12C). In methanol fixed cells the antibody stained the inclusion membrane, however the morphology was aberrant (Figure 7.12D). All further images were from PFA-fixed cells, blocked with 10 % BSA and stained with anti-CT442 at a dilution of 1/200. The rabbit pre-immune sera showed no discernible staining pattern in uninfected or infected HeLa cells.

To confirm the antibody was specifically binding to CT442, HeLa cells previously transfected with the CT442-GFP construct were stained. In uninfected cells anti-CT442 staining closely matched GFP, while also supporting the existence of CT442-GFP derived vesicular structures in the host cell cytosol (Figure 7.13 parts A to C). In infected cells anti-CT442 staining matched GFP but the inclusion membrane was also stained similar to that seen with the antibody in cells without transfection (Figure 7.13 parts D to F). This demonstrated that the anti-CT442 antibody was specifically and consistently staining exogenous and endogenous CT442.

Temporal expression of CT442 throughout the Ct developmental cycle was investigated. No staining was observed at 2, 4, 6 or 12 HPI. Staining of the inclusion membrane was observed from 24 HPI and persisted at 48 and 72 HPI (Figure 7.14), this matched its mid-late cycle gene expression profile. Anti-CT442 uniformly stained the inclusion membrane and occasional fibrous-like extensions could be seen extending from the membrane (Figure 7.14 parts C and F).

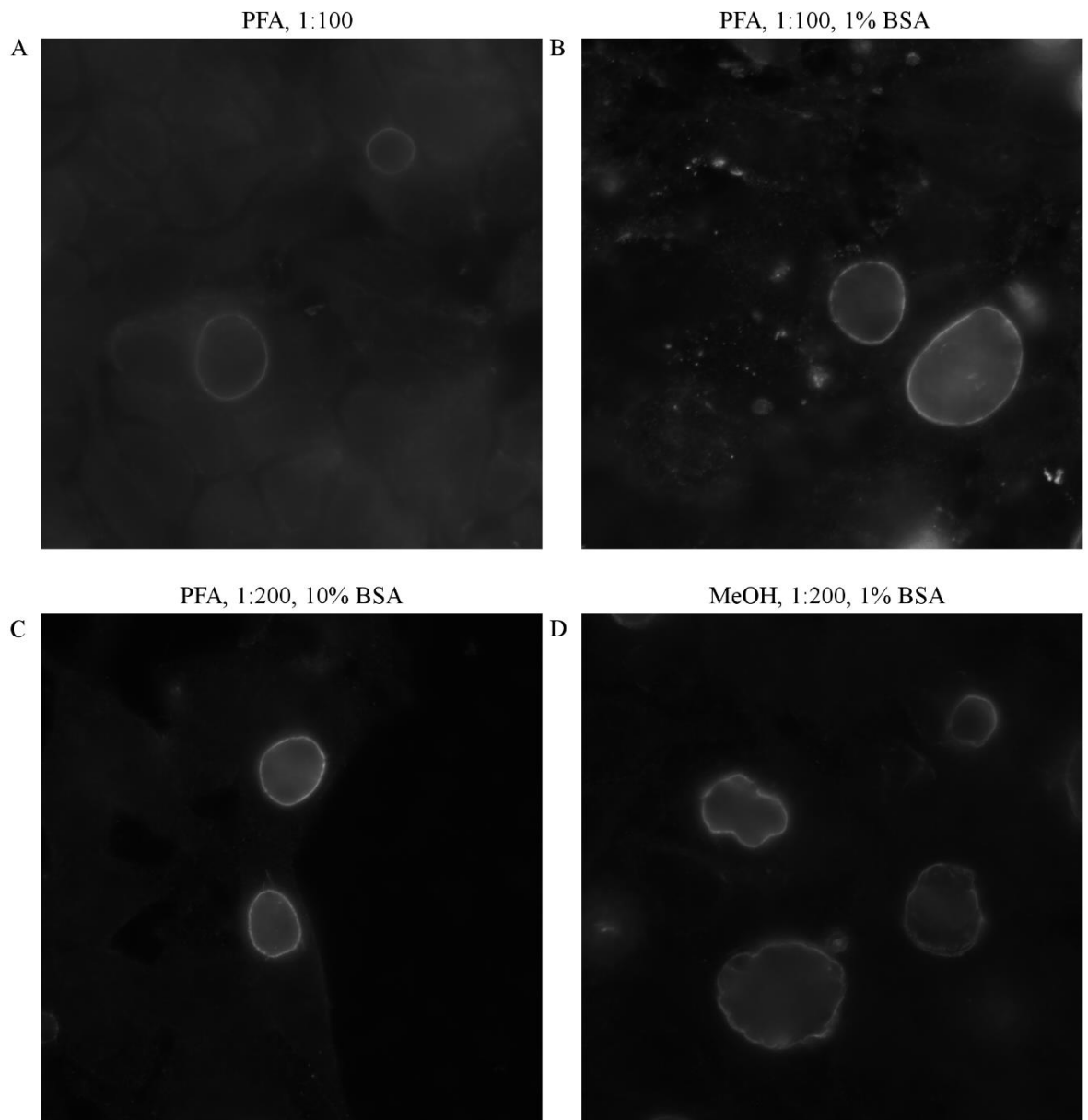


Figure 7.12. Anti-CT442 antibody staining of the inclusion membrane in Ct-L2 infected HeLa cells.

HeLa cells were infected with Ct-L2 and fixed 24 HPI with 4 % PFA (A-C) or methanol (D). Fixed cells were incubated with anti-CT442 antibody at a dilution of 1/100 (A and B) or 1/200 (C and D) after no blocking (A) or blocking with 1 % (B and D) or 10 % BSA (C). Anti-CT442 staining was confined to the inclusion membrane.

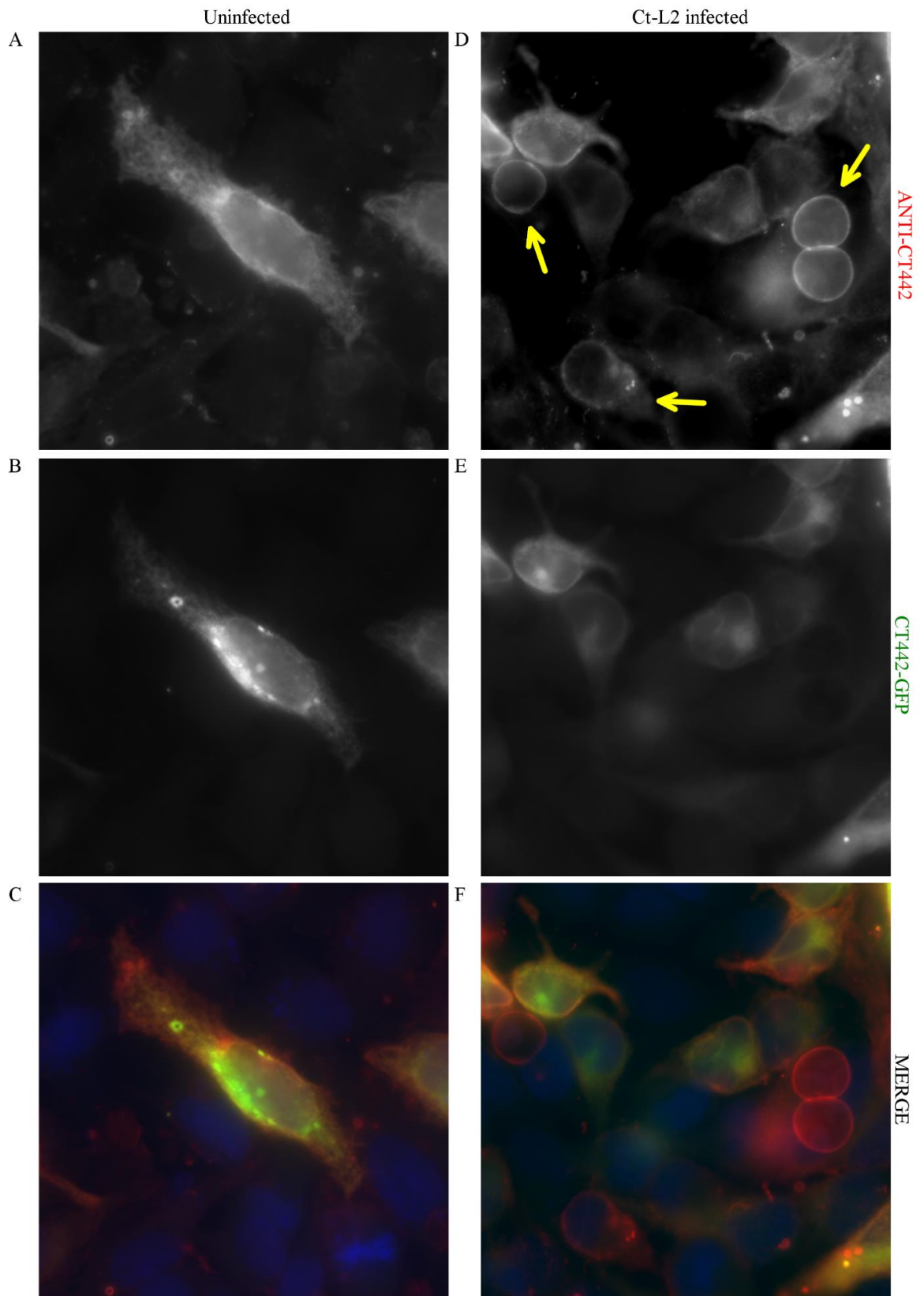


Figure 7.13. Anti-CT442 stains CT442-GFP in HeLa cells.

HeLa cells were transfected with CT442-GFP and fixed after 24 hours (A-C) or then infected with Ct-L2 and fixed 24 HPI (D-F). Cells were stained for DNA (blue in merged panel), CT442 (A, red in merged panel) and GFP (B, green in merged panel). Inclusions are indicated (yellow arrows).

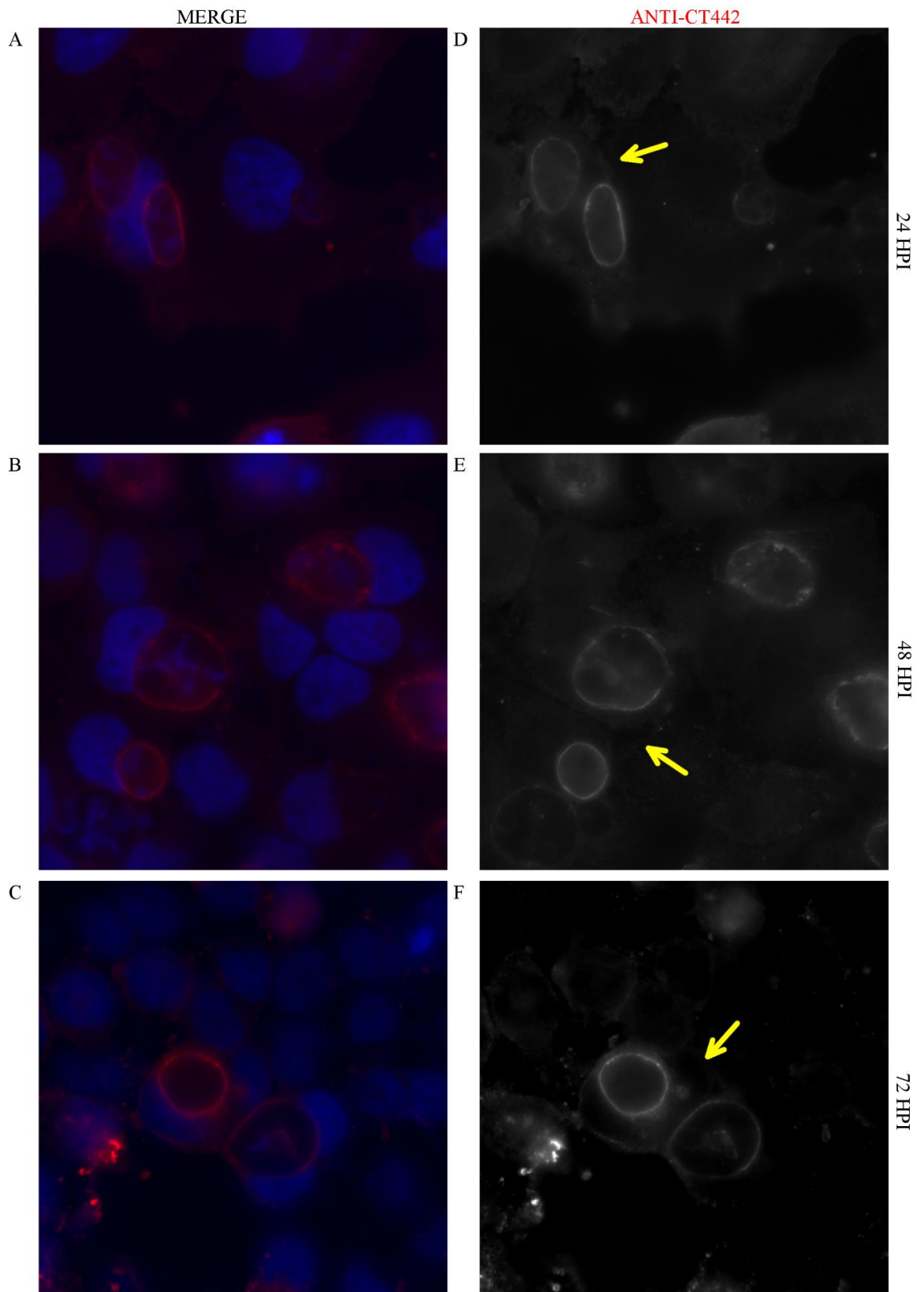


Figure 7.14. Anti-CT442 antibody staining of the inclusion membrane sustains from 24 HPI to 72 HPI.

HeLa cells were infected with Ct-L2 and fixed 24 (A), 48 (B) or 72 (C) HPI with 4 %. Fixed cells were incubated with anti-CT442 antibody at a dilution 1/200 after blocking with 10 % BSA. Anti-CT442 staining was detectable from 24 HPI onwards and remained at 72 HPI. Inclusions are indicated (yellow arrows).

7.2.5. Mass-spectrometry of CT442-GFP Co-IP identified potential interactors

Co-immunoprecipitation of CT442-GFP from HeLa cells was used to identify interactors with CT442 and to highlight particular pathways or cellular functions. CT442-GFP and GFP alone were co-immunoprecipitated from uninfected and Ct infected HeLa cells 48 hours after infection and bound proteins were characterised by mass-spectrometry (MS). Successful purification of CT442-GFP and GFP alone was confirmed prior to MS of the samples (Figure 7.15).

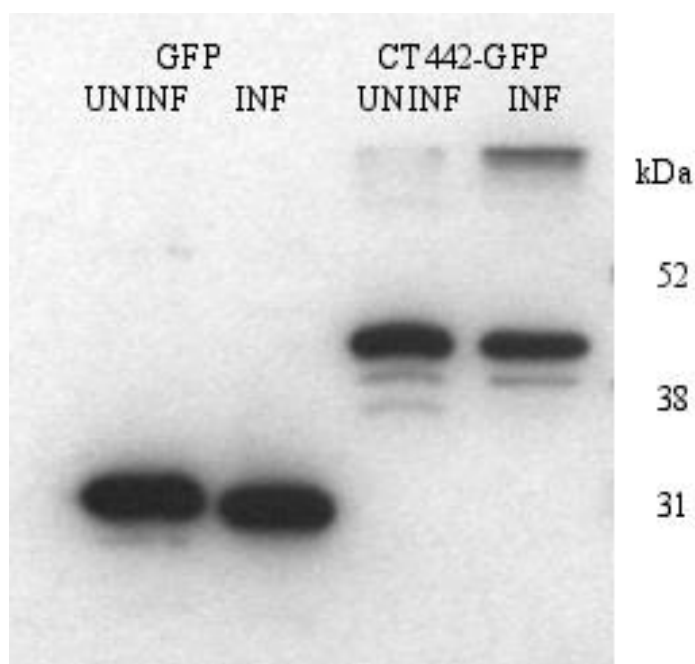


Figure 7.15. Confirmation of GFP and CT442-GFP co-immunoprecipitation.

Western blots were incubated with anti-CT442 antibody to bind CT442. Samples from co-immunoprecipitations of GFP and CT442-GFP transfected HeLa cells were used to confirm the specificity of the antibody for CT442. The experiment was performed for uninfected (UNINF) and Ct-L2 infected (INF) HeLa cells.

Filtering of proteins identified from mass-spectrometry is detailed in chapter 3.7.6 (Table 7.1 and Figure 7.16). Briefly, an initial list of 1347 proteins was reduced to 291 by excluding all hits found with GFP. Keeping those found in CT442-GFP transfected and Ct infected cells and those found in CT442-GFP transfected cells both with and without Ct infection, 127 and 164 respectively. Proteins with scores below 50 were excluded to remove infrequently found proteins leaving 75 filtered targets (Tables 7.2

and 7.3). Forty-nine out of 75 of these were immunoprecipitated by less than five other Incs in a study of the Inc-interactome, this threshold was based on the previously demonstrated presence of 4 Incs together in micro-domains^{97, 376}. Sixteen hits overlapped between these and CT442 results from the published Inc-interactome, only one target was a strongly identified in both and immunoprecipitated by less than five other Incs. This was HYOU1, hypoxia upregulated 1, an ER-localised heat-shock protein involved in protein folding and secretion and protecting cells from hypoxia-induced apoptosis. Additionally, 28/75 proteins were enriched in samples of purified Ct inclusions and 8/75 were enriched in lipid droplets purified from Ct infected cells.

Table 7.1. Grouping of proteins identified by mass spectrometry.

Proteins were identified from uninfected (UNINFECTED) or Ct-L2 infected (INFECTED) HeLa cells by co-immunoprecipitation of GFP or CT442-GFP. A '+' indicated presence of a protein in the individual co-immunoprecipitations. These groupings were used in filtering of proteins described in Figure 7.16.

GROUP	UNINFECTED		INFECTED		DESCRIPTION
	GFP	CT442-GFP	GFP	CT442-GFP	
1	+	-	+	-	GFP only
2	+	+	+	+	Ubiquitous
3	+	+	-	-	Uninfected only
4	-	-	+	+	Infected only
5	-	+	-	-	CT442-uninfected
6	-	+	-	+	CT442-ubiquitous
7	-	-	-	+	CT442-infected

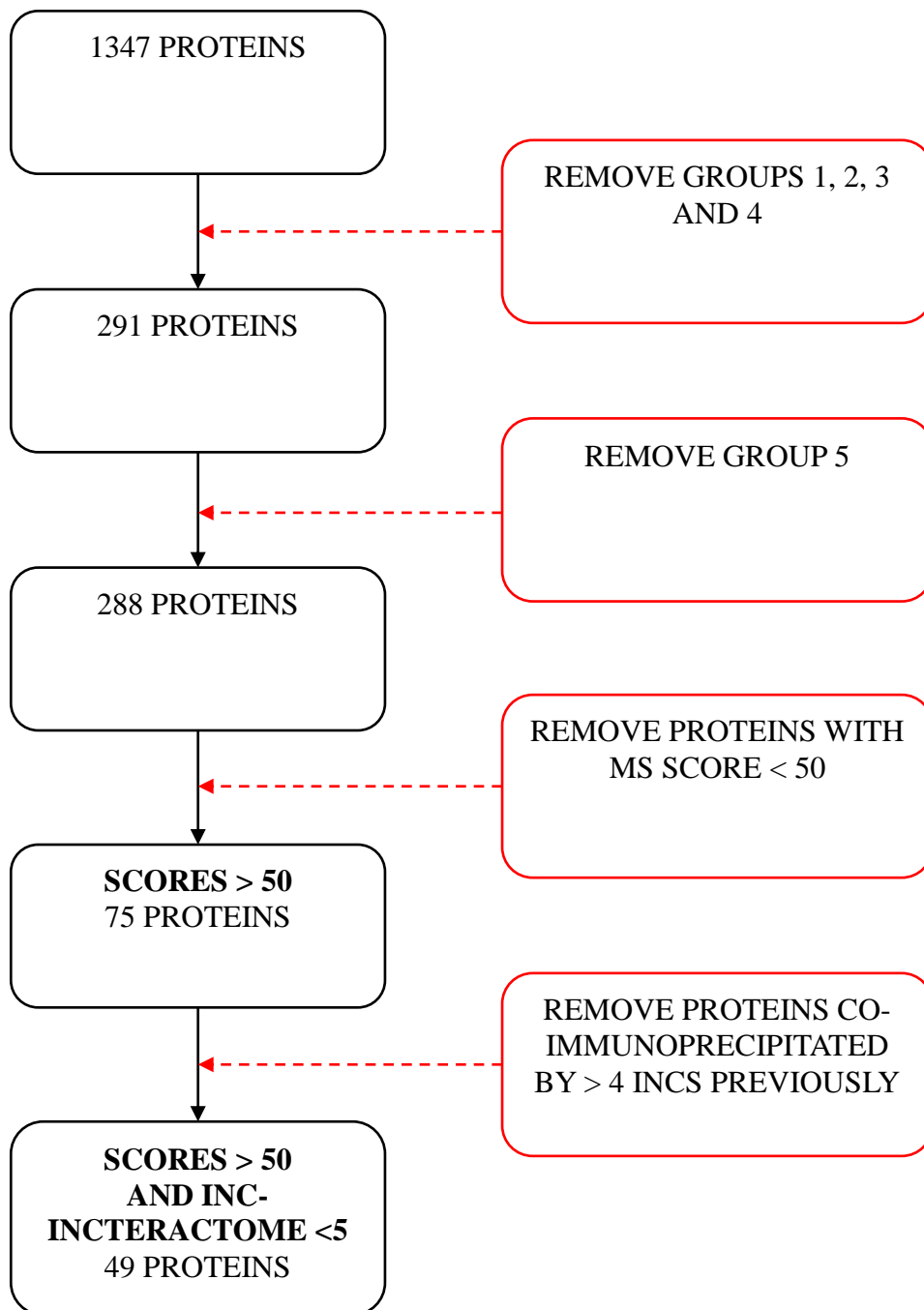


Figure 7.16. Filtering of proteins identified by mass-spectrometry from co-immunoprecipitation of GFP or CT442-GFP.

Black boxes indicate number of proteins after each stage of filtering. Red boxes indicate how proteins were filtered at each step. The two sections in bold are the lists of proteins used for subsequent analyses.

Table 7.2. Summary of number of proteins identified after filtering.

Number of proteins and scores from mass spectrometry of CT442-GFP co-immunoprecipitated from Ct-L2 infected HeLa cells.

	CT442-GFP UNINFECTED	CT442-GFP CT-LGV2 INFECTED
NUMBER OF HUMAN PROTEINS	876	855
MEDIAN SCORE (IQR)	46.04 (34.96-91.49)	43.76 (29.86-73.99)
NOT FOUND WITH GFP	318	290
NOT FOUND IN CT442 UNINFECTED	NA	217
NUMBER OF CT PROTEINS	2	6
MEDIAN SCORE (IQR)	247.48 (152.28-342.67)	35.34 (34.76-53.97)
NOT FOUND WITH GFP	2	4
NOT FOUND IN CT442 UNINFECTED	NA	5

Table 7.3. Details of filtered proteins with a mass spectrometry score greater than 50.

Proteins are identified by their UNIPROT id (ID). The number of times a protein was identified in the host-Inc interactome described previously was indicated (INC-INTERACTOME)⁹⁶. The presence of a proteins in proteomic analysis of Ct inclusions (INCLUSION-ENRICHED)²⁵⁸ and inclusion-associated lipid droplets was indicated (LIPID-DROPLET ENRICHED)⁹³. Function was assigned manually based on UNIPROT descriptions and a PubMed literature search (FUNCTION).

ID	INFECTED SCORE	UNINFECTED SCORE	INC- INTERACTOME	INCLUSION- ENRICHED	LIPID DROPLET- ENRICHED	FUNCTION
PPIA	158.23	84.72	60		Yes	Protein folding
GAPDH	137.97	80.62	62	Yes	Yes	Glycolysis and nuclear functions
RAB7A	120.02	105.31	4		Yes	Endo-lysosomal trafficking
PDCD6	118.95	21.58	0			Multivesicular body pathway
CANX	95.23	178.14	39	Yes	Yes	Protein assembly
EIF3	94.11		36			Part of eIF-3 complex, protein synthesis initiation
VDAC1	92.35	22.43	14			Membrane channel
SRPRB	90.11	74.61	12	Yes		Part of SRP receptor, protein targeting
DAD1	87.48	100.06	5	Yes		Part of N-OST complex, glycosylation
S100A11	84.73	100.81	7			Regulation of differentiation and cell cycle progression
NCLN	79.95	53.39	1			Part of Nodal-modulator complex
GLRX3	79.66	41.91	0			Regulation of cellular iron homeostasis

UCK2	75.61	39.32	2		Nucleotide phosphorylation
MRPL14	73.94		4		Part of ribosome
CYB5R3	72.06	125.53	1		Fatty acid and cholesterol biosynthesis
RAB18	71.40	55.61	6	Yes	Endocytosis and recycling
CLIC1	71.29	89.22	0		Chloride ion channel
SPTLC1	70.57		22		Part of SPT complex
FSCN1	69.26	57.74	0	Yes	Filamentous actin bundling
ESYT2	68.51	77.10	13		Endocytosis and lipid transport
PACSIN2	67.30	45.57	0		Vesicle-mediated transport
RTN4	66.91	121.56	3	Yes	Membrane trafficking
LMAN1	63.73	79.04	0	Yes	Sorting and recycling of proteins and lipids
SCAMP3	63.04	33.01	8	Yes	Post-Golgi recycling
RBBP4	63.03		49		Chromatin regulation
PTGES	60.87	49.99	0	Yes	Prostaglandin synthesis
SORD	60.14	58.39	0		Carbohydrate metabolism
LPGAT1	59.09	76.73	1		Lipid biosynthesis
ITGB1	58.76	87.76	0		Membrane receptor
WDR1	58.75	20.50	0		Actin organisation
GPX8	58.68	37.94	0		Regulates oxidative damage
EMC1	57.67	62.14	3	Yes	Protein folding
NAP1L1	56.83	24.38	37		Chromatin regulation

TMEM9	55.20	50.79	13		Intracellular transport
RAB31	55.18	42.81	0	Yes	Trans-Golgi network (TGN) trafficking
PSMD12	54.41		44		Part of 26S proteasome
CLPTM1L	54.31	100.24	1		Apoptosis
SHMT2	53.82	74.23	3		Amino acid biosynthesis
YIPF5	53.31	60.56	13		ER-Golgi transport
HYOU1	53.08	110.98	3	Yes	Protein folding and responses to hypoxia
ELOVL1	52.01	89.11	2	Yes	Fatty acid synthesis
TMED10	51.42	74.26	15	Yes	Vesicular protein trafficking
RAB13	51.27	42.59	4	Yes	TGN trafficking and tight junction assembly
AGPS	51.06	26.86	5		Phospholipid biosynthesis
PGRMC2	50.16		17	Yes	Steroid receptor
HMOX2	47.66	60.81	11	Yes	Cellular iron ion homeostasis
AGPAT1	46.45	75.86	0		Phospholipid biosynthesis
ALG1	45.70	100.95	2	Yes	Protein glycosylation
JUP	44.05	62.75	13	Yes	Cytoskeleton organisation and function
UBAC2	42.32	52.47	10		ER-associated degradation
POR	42.31	161.55	0	Yes	Electron transfer
MGST1	42.03	73.65	2		Glutathione transferase
ALDH3A2	41.58	66.14	11	Yes	Fatty acid synthesis
KPNA6	39.85	52.37	1		Nuclear protein import

ACTR2	39.61	108.4	5	Yes	Part of ARP2/3 complex, regulates actin polymerisation
ARAP1	39.46	54.03	0		Actin remodelling
GALNT2	38.80	93.06	0		Protein glycosylation
TMCO1	38.45	52.10	3		ER-membrane channel
RAB11A	38.28	78.57	2		Intracellular trafficking
C4A	37.76	60.90	0		Part of complement pathway
ATP13A1	35.92	63.58	0	Yes	Manganese homeostasis
CD97	33.76	92.54	0		Transmembrane signalling
FDPS	32.59	69.96	0		Cholesterol biosynthesis
PTDSS1	30.58	57.12	7		Lipid biosynthesis
RAB32	28.46	77.27	0	Yes	Regulates mitochondria-associated membrane (MAM) properties
USP5	28.28	63.88	0		Ubiquitin disassembly
MBOAT7	27.68	67.56	4	Yes	Phospholipid biosynthesis
SEC22B	26.24	89.32	11	Yes	ER-Golgi and Golgi-derived retrograde transport
DHCR7	25.98	59.68	21	Yes	Cholesterol biosynthesis
TMX1	25.44	54.07	27	Yes	Redox homeostasis
ACADVL	24.33	71.42	0		Fatty acid synthesis
RAB2A	23.22	59.94	3	Yes	Yes ER-Golgi transport
RDH11	22.88	116.90	1	Yes	Retinol metabolism

VASP	20.58	57.19	0	Cell motility
TNPO1	15.97	52.08	17	Nuclear protein import

7.2.6. Pathway enrichment and network analyses highlight intracellular vesicular trafficking

Interactions within the filtered targets were analysed to identify over-represented functions or pathways associated with CT442-GFP, quoted p-values were Bonferroni corrected results from χ^2 tests comparing the filtered list with the complete human proteome. Analyses were done in duplicate using 75 filtered targets and 49 further filtered by excluding proteins found with five or more Incs. These two lists will be referred to as 'Scores > 50' and 'Scores > 50 and Inc-interactome <5'.

Gene Ontology-term biological process analysis highlighted intracellular vesicle trafficking, particularly trafficking involved Rab-proteins, which was supported by GO-term function analysis (Table 7.4). GTPase activity linked with Rab-proteins was also highlighted by protein domain analysis through Pfam and InterPro (Table 7.5). The lists containing 79 and 45 targets were comparable in these analyses.

Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis highlighted metabolic pathways, protein trafficking related to endocytosis and protein processing (Table 7.6). Protein processing was also identified through Reactome analysis, this database only included information for 60/75 and 42/49 proteins respectively which may explain the limited functions identified. The 2 lists of targets were broadly comparable, except for Reactome which failed to identify any over-represented pathways.

Table 7.4.GO-term process and function enrichment of Rab proteins and protein transport.

GO METRIC	SCORES > 50				SCORES > 50 & INC-INTERACTOME < 5			
	TERM	DESCRIPTION	NUMBER	ADJUSTED P-VALUE	TERM	DESCRIPTION	NUMBER	ADJUSTED P-VALUE
PROCESS	GO.0032482	Rab protein signal transduction	7	< 0.001	GO.0032482	Rab protein signal transduction	7	< 0.001
	GO.0016192	vesicle-mediated transport	16	0.006	GO.0007264	small GTPase mediated signal transduction	10	0.007
	GO.0006886	intracellular protein transport	12	0.024	GO.0030036	actin cytoskeleton organization	8	0.007
	GO.0015031	protein transport	15	0.034				
	GO.0007264	small GTPase mediated signal transduction	11	0.036				
FUNCTION	GO.0016491	oxidoreductase activity	13	0.002	GO.0003924	GTPase activity	7	0.005
	GO.0003924	GTPase activity	7	0.033	GO.0043168	anion binding	19	0.005
	GO.0005525	GTP binding	8	0.033	GO.0019003	GDP binding	4	0.008

	GO.0019003	GDP binding	4	0.033	GO.0043167	ion binding	28	0.017
	GO.0043168	anion binding	22	0.033	GO.0005525	GTP binding	7	0.018

Table 7.5. Protein domain enrichment of GTPases.

DATABASE		SCORES > 50			SCORES > 50 & INC-INTERACTOME < 5			
	DOMAIN	DESCRIPTION	NUMBER	ADJUSTED P-VALUE	DOMAIN	DESCRIPTION	NUMBER	ADJUSTED P-VALUE
PFAM	PF00071	Ras family	6	0.007	PF00071	Ras family	6	< 0.001
INTERPRO	IPR003579	Small GTPase superfamily, Rab type	6	0.005	IPR003579	Small GTPase superfamily, Rab type	6	< 0.001
	IPR001806	Small GTPase superfamily	6	0.008	IPR001806	Small GTPase superfamily	6	< 0.001
	IPR002041	Ran GTPase	4	0.012	IPR005225	Small GTP- binding protein domain	6	0.001
	IPR005225	Small GTP- binding protein domain	6	0.012	IPR002041	Ran GTPase	4	0.002
					IPR003578	Small GTPase superfamily, Rho type	4	0.014

Table 7.6. Pathway analysis enrichment of metabolic pathways and processes related to intracellular trafficking.

The two pathways identified by Reactome were involved in the unfolded protein response related to the ER.

DATABASE	SCORES > 50			SCORES > 50 & INC-INTERACTOME < 5		
	PATHWAY	NUMBER	ADJUSTED P-VALUE	PATHWAY	NUMBER	ADJUSTED P-VALUE
KEGG	Metabolic pathways	17	< 0.001	Metabolic pathways	10	< 0.001
	Glycerophospholipid metabolism	4	0.001	Endocytosis	4	0.002
	Phagosome	4	0.010	Glycerophospholipid metabolism	3	0.002
	Protein processing in ER	4	0.010	Leukocyte transendothelial migration	2	0.048
	Endocytosis	4	0.016			
REACTOME	XBPI(S) activates chaperone genes	6	0.028			
	IRE1alpha activates chaperones	6	0.028			

Pathway analyses were supported by looking for interacting partners within the two lists using the Search Tool for Recurring Instances of Neighbouring Genes (STRING). STRING identified 50 and 15 interactions respectively which were both higher than expected p-values of 0.003 and < 0.001. Thirty-five and 9 of these had interactions with scores greater than 0.5 (Table 7.7). Interactions involving Rab proteins, specifically Rab7A, were the highest alongside other proteins involved in intracellular trafficking between the ER and Golgi. In the shorter list filtered using results from the host-Inc interactome, interactions related to glutathione metabolism were also prevalent.

The network analyses for the list of 75 targets focussed around calnexin (CANX) and GAPDH. It involved ER-related trafficking proteins including Rab2, 7, 11 and 13 and less closely those involved in glutathione metabolism (Figure 7.17). In this figure the size of the nodes reflects the score of the individual hit from the MS and the thickness of edges reflects the strength of the interaction indicated by STRING. Both calnexin and GAPDH were not in the list of 49, because they were common in MS results from the Inc-interactome. The network from this further filtered list lacked a clear node, 2 strong routes of interaction remained (Figure 7.18). Three proteins involved in glutathione metabolism were linked and most prominently 4 Rab proteins, including Rab31 whose interaction score with Rab7A was less than 0.5, and ARAP1 which are all involved in intracellular trafficking.

Table 7.7. Ranked protein-protein interactions identified by STRING.

Scores > 50			Scores > 50 & Inc-interactome < 5		
SCORE			SCORE		
RAB7A	RAB11A	0.98	RAB7A	RAB11A	0.98
RAB2A	GAPDH	0.97	RAB13	RAB11A	0.92
RAB13	RAB11A	0.92	GPX8	MGST1	0.90
YIPF5	RAB11A	0.90	ITGB1	VASP	0.84
GPX8	MGST1	0.90	PDCD6	ALG1	0.76
LMAN1	CANX	0.89	ARAP1	RAB13	0.75
ITGB1	CANX	0.84	LPGAT1	AGPAT1	0.70

ITGB1	VASP	0.84	LPGAT1	NCLN	0.64
RBBP4	NAP1L1	0.83	FDPS	SORD	0.54
HYOU1	CANX	0.83			
SHMT2	SPTLC1	0.81			
SHMT2	GAPDH	0.80			
VDAC1	CANX	0.76			
PDCD6	ALG1	0.76			
ARAP1	RAB13	0.75			
LPGAT1	AGPAT1	0.70			
PPIA	GAPDH	0.69			
TMX1	CANX	0.64			
LPGAT1	NCLN	0.64			
VDAC1	GAPDH	0.64			
ELOVL1	SPTLC1	0.63			
TMED10	DAD1	0.63			
RAB13	YIPF5	0.63			
RAB11A	CANX	0.60			
CANX	GAPDH	0.60			
FDPS	DHCR7	0.57			
SHMT2	JUP	0.57			
UCK2	JUP	0.55			
RAB11A	GAPDH	0.55			
ITGB1	GAPDH	0.55			
SRPRB	MRPL14	0.54			
FDPS	SORD	0.54			

TMED10	LMAN1	0.53
--------	-------	------

HYOU1	GAPDH	0.51
-------	-------	------

KPNA6	TNPO1	0.51
-------	-------	------

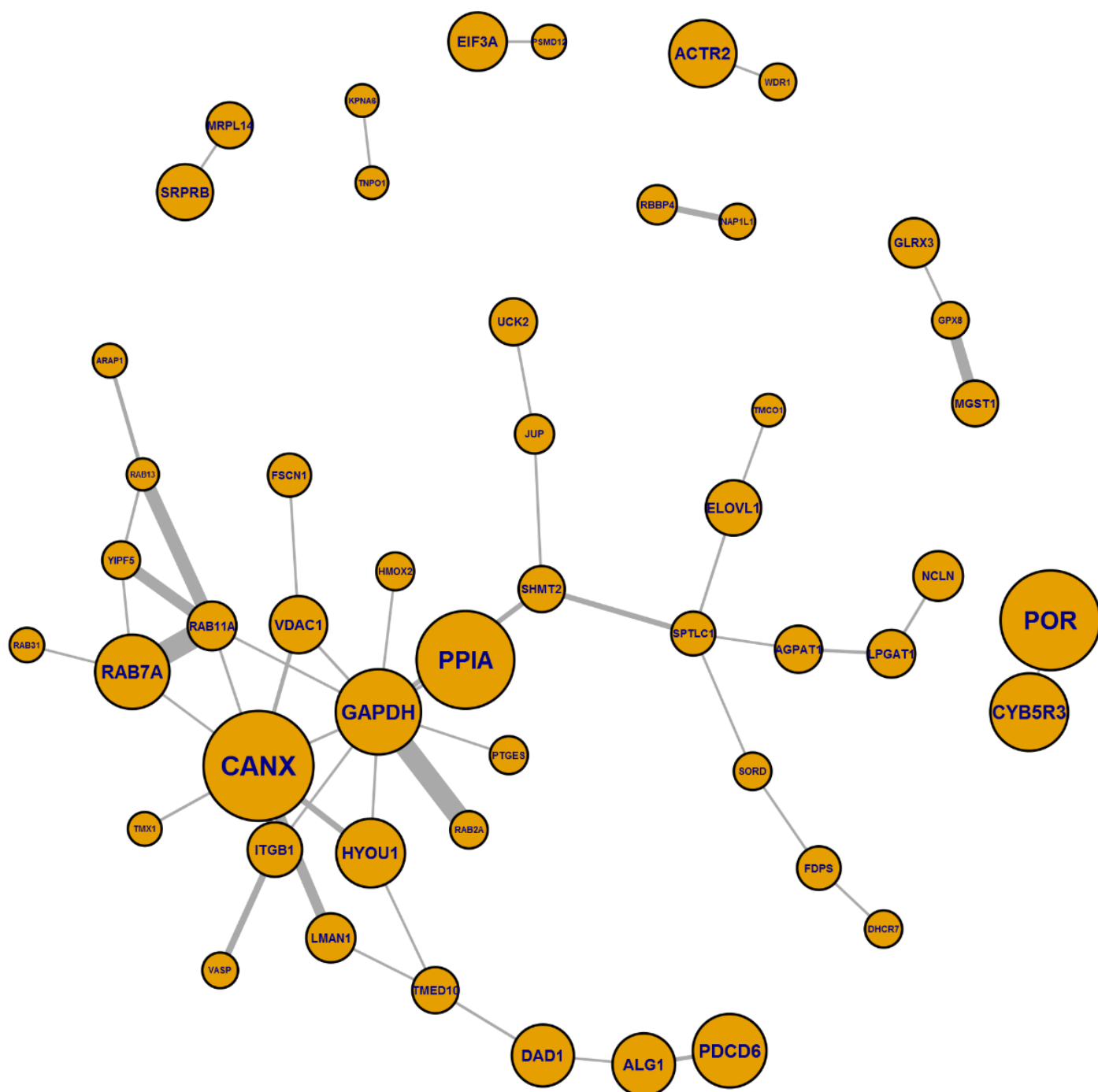


Figure 7.17. A visualisation of protein-protein interactions identified from CT442-GFP co-immunoprecipitation by STRING.

Protein UNIPROT identifiers are indicated within each circle. The size of the circles reflects the proteins score from the mass spectrometry, the larger the circle the higher the score. The thickness of the edges reflects the strength of the interactions identified by STRING, the thicker the edge the higher the score. This was based on the list of 75 proteins, including all interactions not just those with a score above 0.5 as in Table 7.6.

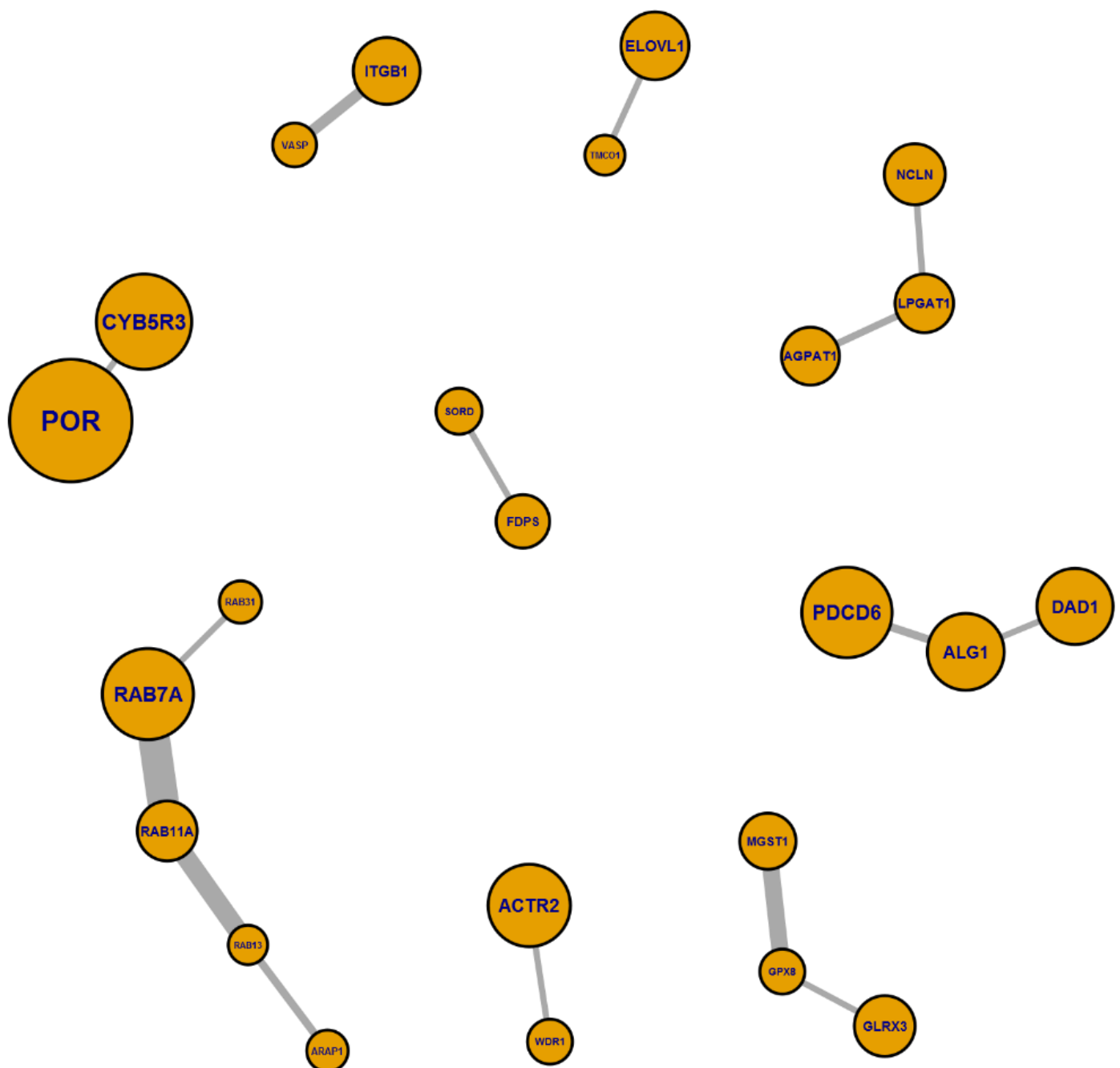


Figure 7.18. A visualisation of protein-protein interactions identified from CT442-GFP co-immunoprecipitation by STRING.

Protein UNIPROT ids are indicated within each circle. The size of the circles reflects the proteins score from the mass spectrometry, the larger the circle the higher the score. The thickness of the edges reflects the strength of the interactions identified by STRING, the thicker the edge the higher the score. This was based on the list of 49 proteins, including all interactions not just those with a score above 0.5 as in Table 7.6.

7.2.7. Chlamydial targets link with human pathway enrichment of vesicular trafficking

Six hits were found in the NCBI Ct protein database, after removing CT442 and those found with GFP or in uninfected cells 3 remained (Table 7.8). CT049 which is inclusion-associated³⁸⁴, CT147 which is an Inc and CT853 which resides in the Ct inner membrane. Functions of these 3 were not known, however, CT147 is a homologue of early endosomal antigen 1 (EEA1)²¹², a eukaryotic protein involved in endosomal trafficking known to interact with Rab proteins³⁸⁵.

Table 7.8. Ct proteins identified through CT442-GFP co-immunoprecipitation.

Ct D/UW3 nomenclature was used (ID). Localisation and function were assigned manually based on a PubMed literature search.

ID	INFECTED SCORE	LOCALISATION	FUNCTION
CT049	35.68	Inclusion lumen	Pmp-like
CT147	34.99	Inclusion membrane	EEA1 homologue
CT853	34.68	Inner membrane	MarC homologue

7.2.8. Confirmation of Rab7 immunoprecipitated by CT442-GFP

Co-immunoprecipitation of Rab7 by CT442-GFP was selected for investigation due to its strong mass-spectrometry score and key involvement in the highlighted pathways and interactions focussed around Rab proteins and vesicular trafficking.

Specificity of the anti-Rab7 antibody was confirmed by western blotting of HeLa cells lysates (Figure 7.19A). Samples from the co-immunoprecipitation CT442-GFP, blotted previously in Figure 7.15, were examined by western blot for the presence of Rab7. Rab7 was detectable in co-immunoprecipitations of CT442-GFP from uninfected and Ct-infected HeLa cells (Figure 7.19B).

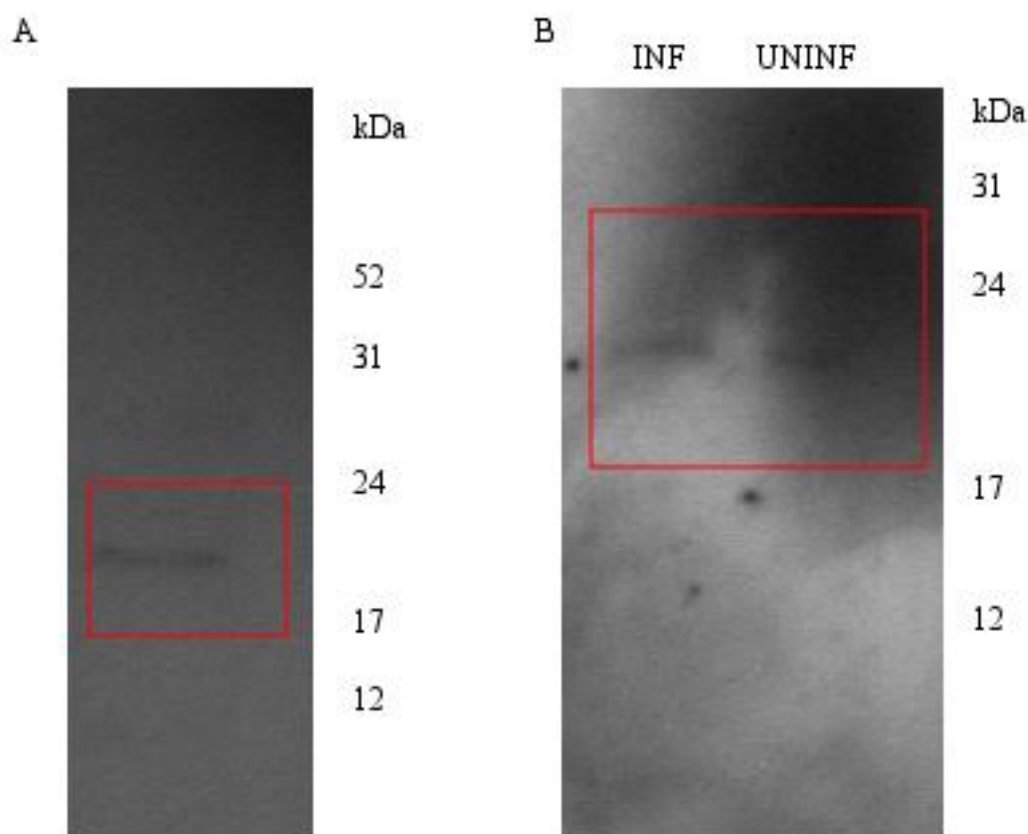


Figure 7.19. Confirmation of Rab7 co-immunoprecipitation with CT442-GFP.

Western blots were incubated with anti-Rab7 antibody to bind Rab7. A) Anti-Rab7 antibody specificity was confirmed in HeLa cell lysates, a band at the expected molecular weight (23 kDa) was detected. B) Rab7 co-immunoprecipitation with CT442-GFP was confirmed in uninfected (UNINF) and Ct-L2 infected (INF) HeLa cells.

7.3. Discussion

In this study cell-culture models were used to examine the intracellular localisation, trafficking and potential interactions of the previously identified immune target CT442. The membranous nature of the protein and its localisation to the Ct inclusion membrane during infection of mammalian cells were confirmed initially using a GFP-tagged CT442 construct and later an affinity-purified monoclonal CT442 antibody. Co-immunoprecipitation using the GFP-tagged construct and subsequent mass spectrometry analysis of bound proteins allowed identification of potential CT442-interactions. Protein domain and pathway analyses found few over-represented pathways but

generally agreed on the presence of proteins involved in varying forms of intracellular trafficking and endosomal cycling. Examining interaction pairs and strings supported this and highlighted Rab7A, amongst other Rab family proteins, as being the most abundant constituent of proteins immunoprecipitated alongside CT442. This was strengthened further by one of only 3 Ct proteins that were identified being CT147, another Inc, which is a homologue of the mammalian protein EEA1 involved in endosomal trafficking.

7.3.1. CT442 is polymorphic and it is not essential for Ct survival

Nucleotide and amino acid sequence analysis of CT442 in the *Chlamydia* genus found conservation of the predicted transmembrane helices, the N-terminal and C-terminal exposed regions were not conserved. This suggests localisation to the inclusion membrane may be a conserved feature, but functional interactions may vary. Within Ct there was polymorphism in the predicted transmembrane helices and the cytosol-exposed termini. This supports the findings from chapter 6 where within a population of ocular Ct isolates CT442 had evidence of selection in three regions, overlapping each termini and the first transmembrane helix respectively. Selection in the cytosol-exposed termini could be explained by immune targeting of these domains or polymorphisms affecting currently undefined functional interactions. There was evidence of both purifying and positive selection within the first transmembrane domain.

Both forms of selection have been observed previously in transmembrane domains, although the pressure driving positive selection has not been examined^{386, 387}. This domain contains an epitope capable of inducing CD8⁺ T-cells in murine Ct infections¹⁹⁵. How Incs are inserted into the inclusion membrane is currently not known. It is possible that between type-3 secretion from within the inclusion lumen and insertion into the inclusion membrane CT442 could be diverted to pathways involved in antigen presentation and subsequently induce T-cell responses.

This study identified a novel truncation mutant of CT442 in an ocular Ct isolate from the Bijagos Island, Guinea-Bissau. This truncated the protein after amino acid 99, considerably later than two previously observed truncation mutations in two serovar C strains. This is the first evidence of a CT442 truncated mutant outside of serovar C. These mutations show that CT442 is not essential for Ct survival. Serovar C strains are

rarely found in trachoma-endemic communities of sub-Saharan Africa, although they have no growth defects *in vitro* despite possessing several other truncated genes as described previously^{382, 383}. It is not known if these truncations and specifically those in CT442 are related to low prevalence.

7.3.2. CT442 localised to the inclusion membrane from 24 hours post infection

CT442 ectopically expressed in mammalian cells by transfection of the CT442-GFP construct localised primarily to the endoplasmic reticulum membrane, confirming the membranous nature of CT442. Anti-CT442 antibody directed against a C-terminal epitope bound this ectopically expressed CT442. This means the termini which *in vivo*, are predicted to be exposed to the cytosol, were oriented this way in this expression system. This was important for interpretations of CT442-interacting proteins identified by co-immunoprecipitation from cells expressing CT442.

In Ct infected HeLa cells anti-CT442 antibody stained the inclusion membrane. This staining was not detectable prior to 24 hours post infection (HPI). Following 24 HPI, staining was detectable up to and including 72 HPI. This confirms transcriptional profiling of CT442, in which expression began at 16 HPI and peaked between 24 and 36 HPI^{212, 213}. This pattern of expression is similar to 8 other Incs previously studied, including IncA and 3 Incs which localise to discrete microdomains^{163, 388}.

Anti-CT442 antibody staining was uniformly distributed around the inclusion, no intra-inclusion or cytosolic staining was seen. This data confirms CT442 is a mid-late cycle expressed Inc, localising exclusively to the inclusion membrane in HeLa cells and not to other microdomains.

7.3.3. Co-immunoprecipitation of CT442-GFP from Ct infected cells identifies intracellular trafficking and Rab proteins

Proteins potentially interacting with CT442 in mammalian cells were identified by co-immunoprecipitation of CT442-GFP from HeLa cells both uninfected and Ct-L2 infected and mass spectrometry analysis of the samples. After filtering out proteins identified in control experiments using GFP alone, 75 proteins were found. Proteins

identified in co-immunoprecipitations of more than 4 Inc from a previous study of Inc-host interactions were excluded, 49 proteins remained. Analysis of the 49 and 75 proteins found mostly similar results.

Rab proteins involved in signal transduction were over-represented in both lists by GO-term process analysis. This was supported by GO-term function which showed enrichment of GTPase activity. Protein domain analysis through Pfam and InterPro identified enrichment of the Rab and Ras families, further strengthening the over-representation of Rab proteins. Pathway enrichment analysis through KEGG and Reactome identified metabolic pathways, endocytosis and protein trafficking. The latter two of these support the importance of Rab proteins.

Protein-protein interactions identified by STRING showed greater discrepancy between the two lists of 49 and 75 proteins. The interactions involving the 75 proteins were centred around CANX (calnexin), GAPDH (glyceraldehyde 3-phosphate dehydrogenase) and PPIA (peptidylprolyl isomerase A). These were identified in a high number of Inc interactors. They are probably artefacts of the co-immunoprecipitation not genuine CT442 interactors. The interactions involving the 49 proteins had two strings with more than two proteins with an interaction score greater than 0.5. One involved NCLN (nicalin precursor), LPGAT1 (acyl-CoA:lysophosphatidylglycerol acyltransferase 1) and AGPAT1 (1-acylglycerol-3-phosphate O-acyltransferase 1). These interactions are indicative of metabolic pathways highlighted by KEGG. The other involved Rab7, Rab11, Rab13 and ARAP1 (ArfGAP With RhoGAP Domain, Ankyrin Repeat And PH Domain 1). These interactions were related to the Rab proteins and protein trafficking pathways highlighted previously.

Ct proteins identified in the mass spectrometry supported the enrichment of intracellular trafficking pathways and Rab proteins. Two of the 3 identified Ct proteins localise within the inclusion, the other was an Inc CT147. CT147 is a homologue of human early endosomal antigen 1 (EEA1). This protein is involved in endosomal trafficking and fusion in conjunction with Rab5^{385, 389}. CT147 is homologous in the regions involved in endosome tethering but lacks homology in the regions for endosomal fusion. This finding and expression as early as 1-hour post infection have led to speculation that CT147 is involved in the non-fusogenic properties of Ct²¹². The function of CT147 is unknown but co-immunoprecipitation of this protein with CT442-GFP supports the involvement of intracellular vesicular trafficking and Rab proteins with CT442's potential functions. Co-immunoprecipitation of Rab7 by CT442-GFP was

confirmed by western blot an anti-Rab7 antibody in both uninfected and Ct-L2 infected HeLa cells.

7.3.4. Rab proteins and *Chlamydia trachomatis*

Rab proteins are a large family of GTPases which function in all forms of membrane trafficking within cells, more than 60 mammalian Rab-GTPases have been identified³⁹⁰. Through interactions with a plethora of host factors these proteins control endosomal trafficking between a number of host organelles including; the plasma membrane and early endosomes (Rab4 and Rab5), early and late endosomes (Rab7), the ER and Golgi (Rab1 and Rab6) and endosomes and Golgi (Rab9 and Rab11)^{390, 391}. Ct exploits host membrane trafficking pathways for initial invasion and avoidance of lysosomal fusion and for recruitment of organelles required for intracellular development and survival^{73, 392}. Therefore, Ct likely interacts extensively with Rab proteins.

Preliminary investigations of Rab proteins in Ct identified Rab 1, 4, 6 and 11 enriched at the periphery of the inclusion⁹⁰. This localisation was observed as early as 1 HPI. Since then several Rab proteins have been studied independently in Ct infection. Rab4 was found co-localised with the Inc CT229 at 2 HPI³⁹³. Rab11 partially co-localised with IncG⁹⁰, which has been implicated in Ct-resistance to apoptosis via sequestration of pro-apoptotic BAD through 14-3-3 β ⁹⁴. Rab11 and also Rab6 also function in Ct-driven Golgi fragmentation and related nutrient acquisition⁹¹. Similarly, the Golgi-associated Rab14 is involved in delivery of sphingolipids to the inclusion³⁹⁴. Initial examination of Rab7 in Ct infected cells did not show the inclusion-peripheral staining described for other Rab proteins, however Rab7-staining did appear to cluster in one region of the cells⁹⁰. It is possible that Rab7 has a role in Ct infection without being directly recruited to the inclusion.

Rab7 is primarily associated with late endosomes. It was first identified in the transition from early endosomes to late endosomes, which occurs through the downregulation of Rab5 and upregulation of Rab7³⁹⁵. However it has now known to function in retrograde trafficking between endosomes and the Golgi³⁹⁶, recruitment of the proteasome and nutrient transport through lipid trafficking. Rab7 also regulates phagosomal maturation post-phagocytosis. For this reason, intracellular pathogens have developed a number of strategies to modulate Rab7 and therefore phagosomal

maturation. These can be broadly grouped into two strategies³⁹⁷. Pathogens can downregulate Rab7 to prevent recruitment to and subsequent maturation of phagosomes, utilised by *Mycobacterium* and *Leishmania donovani*. Alternatively, pathogens can partially recruit Rab7 leading to formation of a compartment that is both late endosomal and phagosomal but crucially not lysosomal, utilised by *Salmonella enterica* and *Helicobacter pylori*. Ct infection of macrophages leads to suppressed growth due to lysosomal targeting, this follows rapid association of intracellular Ct with Rab7 and LAMP1, another marker of late endosomes³⁹⁸. Avoidance of lysosomal degradation by Ct in epithelial cells therefore likely requires some form of Rab7 modulation.

Rab11, Rab13 and ARAP1 were also co-immunoprecipitated by CT442-GFP, and identified by STRING as interacting with one another. Rab11 and Rab13 are involved in recycling endosomes from the Golgi to the plasma membrane, Ct does not fuse with recycling endosomes but they do associate with the inclusion^{90, 399}. ARAP1 regulates endocytosis and trafficking of epidermal growth factor receptor (EGFR), which is involved in Ct attachment and inclusion development⁴⁰⁰. ARAP1 interactions precede the association of EGFR with EEA1-positive early endosomes and prevent its degradation⁴⁰¹, which downstream is regulated by Rab7⁴⁰². It is possible CT442 is involved in the recruitment of recycling endosomes via Rab11/13 or EGFR-positive endosomes through ARAP1, currently this is unknown.

7.3.5. Lipid droplets and *Chlamydia trachomatis*

Lipid droplets are host organelles primarily involved in lipid storage. Recent studies have demonstrated they are recruited to the inclusion and can be translocated into the inclusion lumen⁸⁶. The mechanisms and importance of this are unknown, however enrichment of lipid metabolism in these organelles and recruitment to the inclusion periphery was essential for Ct replication⁹³.

Apart from the previously described intracellular trafficking, lipid metabolism was the only other commonly over-represented pathway in proteins potentially interacting with CT442. Proteomic analysis of lipid droplet during Ct infection of mammalian cells identified four associated IncS, including CT442. Many host proteins were enriched as well, including several Rab proteins such as Rab7, Rab11 and

Rab31⁹³. The association of lipid droplets with the inclusion is first detectable 18 HPI, this matches the mid-late cycle expression of CT442⁸⁶. The interactions which recruit lipid droplets to the inclusion and are involved in lipid acquisition are not clearly defined. It is possible CT442 and the Rab proteins identified here are involved in these processes.

7.4. Conclusions and future work

A currently circulating Ct isolate with a CT442 truncation showed that it is not essential for Ct survival. Evidence of selection acting on this gene within a population of ocular Ct isolates suggested that mutations in CT442 do have an impact on Ct survival and transmission. CT442 was confirmed to be a mid-late cycle expressed inclusion membrane protein. Co-immunoprecipitation of CT442-GFP from mammalian cells highlighted Rab proteins, including Rab7, and others involved in intracellular trafficking as important for the function of CT442. Co-immunoprecipitation of Rab7 by CT442-GFP was confirmed, this work therefore found potential CT442-host interactions which require in depth investigation.

Future work should initially utilise the anti-CT442 antibody to co-immunoprecipitate endogenous CT442 from mammalian cells to validate proteins identified using exogenously expressed CT442 through transfection of CT442-GFP into host cells. Anti-Rab7 antibody should be used to determine if endogenous CT442 also interacts with Rab7. If the interaction is confirmed the localisation of Rab7 should be investigated in Ct infected mammalian cells. Rab7 interactions with other bacteria have been shown to vary considerably with different tissues and species³⁹⁷. Therefore, localisation should be investigated in conjunctival and genital epithelial cell lines with ocular and urogenital serovars, including serovar C isolates expressing a truncated CT442. The impact of expression of a Rab7 dominant negative mutant on Ct infectivity should also be investigated. The reported association of Rab7 and lipid droplets during Ct infection should be investigated in a similar fashion as another potential route of interaction between CT442 and Rab proteins.

8. Discussion

8.1. Natural selection within a population of ocular Ct isolates

In this study 126 ocular Ct isolates collected from trachoma-endemic communities on the Bijagos Islands, Guinea-Bissau were screened for genome-wide signatures of selection. In total, 31 genes had evidence of selection at the gene or epitope-level by a combination of Tajima's D and Fay and Wu's H. Eleven of these were under positive selection, 16 were under purifying selection and 4 had regions under both positive and purifying selection. A further 20 genes had evidence of positive selection determined using integrated haplotype scores.

8.1.1. Selection validates Ct host-interactors

Positive selection acting on genes suggests mutations in them provide Ct with some fitness advantage. Conversely, purifying selection suggests mutations in the relevant genes puts Ct at a disadvantage. These genes would be expected to have important functions in Ct infectivity, survival and transmission as these results found.

Genes expressed very early and very late in the Ct developmental cycle were over-represented in those identified as under selection. Studies of Ct and the host have identified these stages as pivotal in determining the success of Ct infection. The most strongly associated SNP from a genome-wide association study of scarring trachoma fell in PREX2⁴⁰³. The protein encoded by this gene activates Rac1, a signalling protein known to be important for TARP-induced changes in the actin cytoskeleton upon Ct invasion of host cells^{61, 404}. This suggested to the authors that scarring trachoma, which is driven by repeated Ct infection and persistent inflammation, is linked to the ability of Ct to efficiently infect epithelial cells. In this thesis I found evidence of positive selection in translocated actin-recruiting phosphoprotein (TARP) and the inclusion membrane protein (Inc) CT147. TARP is type-3 secreted into the cytosol of host cells upon contact of EBs with the plasma membrane, following this TARP induces reorganisation of the actin cytoskeleton and downstream signalling which underlies Ct invasion^{53, 363}. CT147 is a homologue of the mammalian protein early endosomal antigen 1 (EEA1)²¹². This protein is involved in endosome tethering and fusion, linking early endosomes with the late endosomal and lysosomal pathways³⁸⁵. CT147 is expressed as early as 1-hour post infection and is thought to be important in Ct

avoidance of fusion with the lysosomal and therefore continued intracellular survival. Mutations in these genes likely impact their interactions with host factors, improving efficiency of invasion and early intracellular events. CT147 also had evidence of purifying selection, this could indicate a region of structural importance or where precise functional conservation is important.

In Ct the mode of exit from cells has been linked to improved survival and dissemination⁷². Approximately 76 % and 32 % of Ct EBs which exited cells within membrane-bound extrusions were viable at 4 and 24 hours, compared with 40 % and 3 % for EBs which exited via cell lysis¹⁰¹. These extruded EBs are also reportedly able to survive within macrophages, providing a further advantage for transmission. This indicates that the balance of mechanisms controlling cell exit of Ct are important in continued survival and transmissibility in the host. I found evidence of positive selection in CT228 an Inc. CT228 appears to selectively recruit and activate proteins of the myosin phosphatase pathway⁹⁹. Through downstream activation of different myosin motor proteins this interaction regulates Ct exit, favouring extrusion of cell lysis. The ability of CT228 to recruit these proteins may therefore be directly related to how Ct exits cells and subsequently its ability to reinfect. Mutations in CT228 may be altering the balance of exit by extrusion or cell lysis.

Genes under selection were also over-represented by proteins localised to the outer membrane, inclusion membrane and proteins secreted into the host cytosol. This further supported the enrichment of signatures of selection in Ct factors interacting with the host. Twenty-one secreted proteins had evidence of selection, 20 of which were type-3 secreted effectors. These included CT694 and CT868 (ChlaDub1), both had evidence of positive selection. CT694 is a commonly recognised antibody target in trachoma-endemic communities^{286, 405} and is secreted into host cells by Ct early during invasion⁶³. CT694 associates with the host plasma membrane where it interacts with AHNK⁶³. Through this it is hypothesised to alter actin dynamics, effectively reversing TARP-dependent actin rearrangements⁴⁰⁶. ChlaDub1 is a deubiquitinating and deNeddylating protease⁴⁰⁷. Through binding and stabilising of the inhibitory subunit of NFκB, ChlaDub1 can inhibit induction of numerous pro-inflammatory signals⁴⁰⁸.

Half of the 20 T3S effectors are Incs. Selection in these proteins was variable; 5 had evidence of positive selection, 4 had evidence of purifying selection and CT147 had evidence of both. CT229 had evidence of positive selection. This Inc interacts with Rab4⁸⁹, a recycling-endosome related protein, and is therefore thought to be involved in

the intracellular trafficking of the inclusion. CT116 (IncE) was also under positive selection. IncE recruits the retromer components sorting nexins 5 and 6 to the inclusion⁹⁶. This interruption of retrograde trafficking was demonstrated to enhance Ct progeny formation. CT223 (IPAM) had evidence of purifying selection. IPAM is critical for Ct control of the host microtubule network through interactions with CEP170⁸⁷. This interaction is required for Ct intracellular development and survival. The implication from this is that CT223 cannot sustain sequence mutations without impairment of its essential function.

Four EB outer membrane proteins had evidence of selection. CT396 had evidence of positive selection, CT414 (polymorphic membrane protein [pmp] C), CT681 (*ompA*) and CT872 (*pmpH*) had evidence of purifying selection. The functions of these proteins are not known. It is possible that some of them are structurally important for the EB, which could explain purifying selection. In the evolution of transmembrane domains there appears to be considerable evidence of this form of selection^{386, 387}.

8.1.2. Selection identifies genes coding for immunogenic proteins

The previously mentioned outer membrane proteins shared another common feature, they are all immunogenic in one of or both urogenital and ocular Ct infection^{170, 191, 364, 409}. Fifteen of the 31 genes under selection were known to be immunogenic, supporting immune responses as a key driving factor of selection within this population of ocular Ct isolates. Excluding the outer membrane proteins, these included 6 Incs and 5 secreted proteins.

Considerable work has been undertaken to identify antibody targets in Ct infection, particularly those with associations to clinical outcomes of infection. The abundance of antibodies not targeted to the EB surface is not a new finding²⁸⁷, however almost no work has been done to investigate how antigens not exposed on the surface are targeted by antibodies. Based on the *in vitro* developmental cycle there are three possible routes where intracellular antigens could enter the extracellular space and become exposed to antibodies; membrane vesicles, cell lysis and extrusions.

Membrane vesicles are the most limited of these pathways (Figure 8.1A). They are vesicular bodies derived from the Ct cell envelope³⁰² and after release from the inclusion they have been observed in association with the inclusion, the plasma membrane and extracellularly³⁰³. Proteomic analyses of these vesicles have identified MOMP, Pgp3, CPAF, CT159 and Inc, F and G^{302, 304}. These are all known to be immunogenic while only MOMP is expressed on the EB surface, supporting membrane vesicles as a method for generation of antibodies against intracellular targets. Composition of these vesicles is not entirely uniform, it is possible that further antigens could be presented through this mechanism.

The second and third methods are centred on exit of Ct from infected cells. Cell lysis was previously thought to be the only mechanism for exit by Ct. Lytic exit is marked by complete rupturing of the inclusion followed by the cell membrane (Figure 8.1B)⁷². This theoretically would allow the antibodies to target any protein not contained within the bacteria, including secreted proteins, Incs and proteins found in the inclusion lumen.

Similarly to membrane vesicles, the repertoire of antigens that could be presented by extrusions is likely more limited. Extrusion is defined as the packaged release of EBs from infected cells (Figure 8.1C)⁷². This is an active process requiring actin polymerisation whereby the whole or a portion of an inclusion pinches off from cells into a membranous extrusion. Extrusion appears to be conserved amongst *Chlamydia* species and may facilitate greater survival and transmission of the bacteria¹⁰¹. As described previously, extruded EBs were viable considerably longer post cell exit than EBs released through lysis. They can also promote uptake into and survival within macrophages, potentially allowing increased dissemination. The *in vivo* existence and potential prevalence of extrusion is not known, however they do provide a potential mechanism for Incs to be targeted by antibodies. A recent study of extrusions reported that they have a double membrane structure consisting of an intact layer of plasma membrane with the intact EB-containing inclusion¹⁰¹. Electron microscopy images showed limited evidence of this double membrane and staining of extrusions with IncA and a plasma membrane marker seemed to be indistinguishable, suggesting the membrane may be a hybrid of host and Ct proteins. Furthermore, the demonstrated phagocytic-uptake of extrusions would require non-self recognition, supporting the exposure of Incs on the surface. Incs exposed in this way would be readily targetable by antibodies. Based on the enhanced extracellular survival of extrusions they would also

be exposed for longer than EB surface proteins. This method for intracellular antigens to become antibody targets also provides the clearest hypothesis for how induced antibodies could impact Ct survival. Extrusion-targeted antibodies could bind the surface promoting uptake into immune cells for destruction or epithelial cells for reinfection.

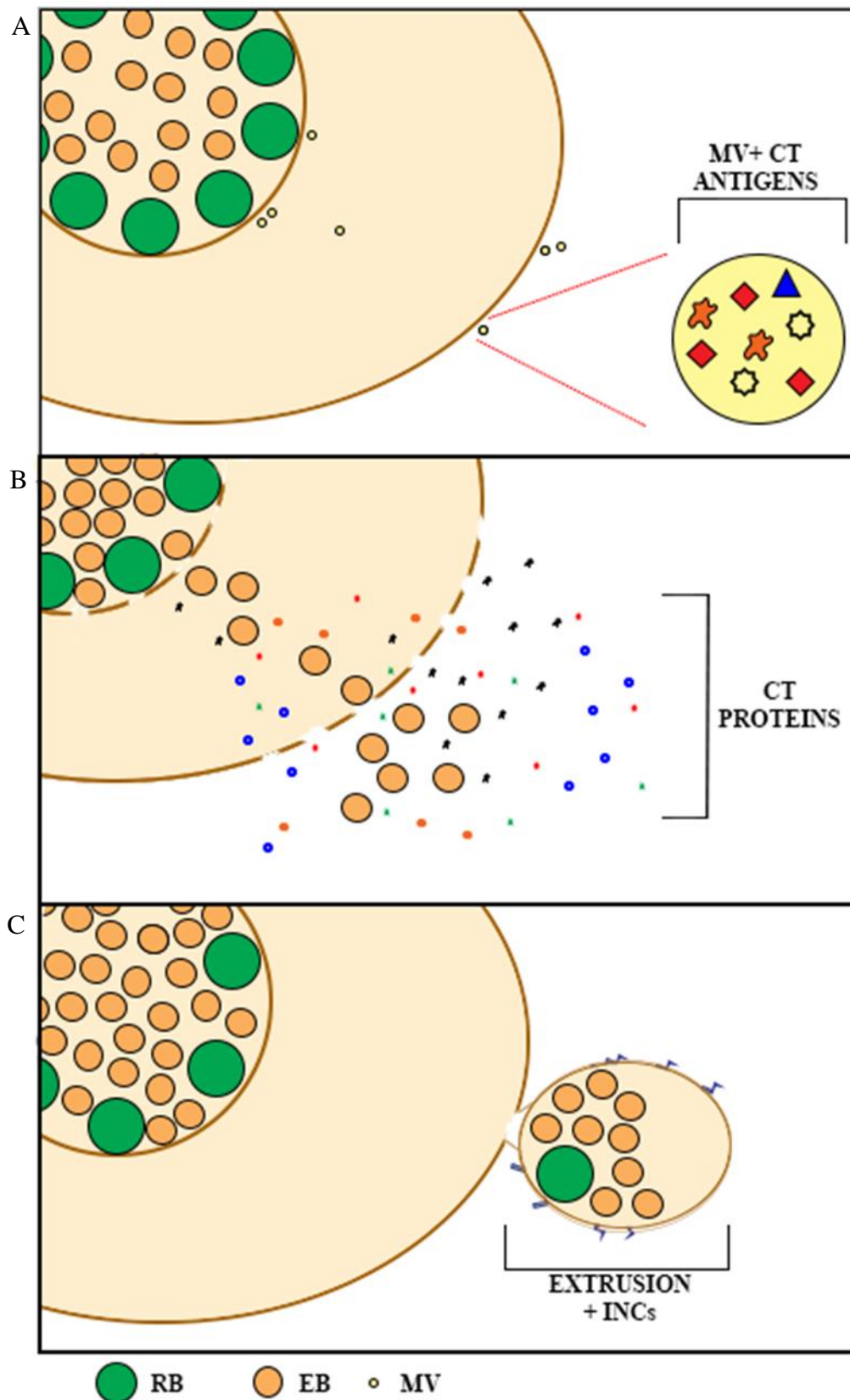


Figure 8.1. Extracellular exposure of non-surface Ct antigens.

Exposure could be achieved through membrane vesicles (MV) releasing select antigens (A), cell lysis releasing many antigens (B) and extrusions presenting Incs (C).

A surprising finding from these genes under selection was a lack of balancing selection, particularly in the immunodominant EB surface proteins. Balancing selection is a common method of immune evasion employed by human pathogens^{343, 350, 367, 368}. Balancing selection maintains multiple allelic forms of immunodominant genes in a population through cyclical presentation of variable forms of antigens, allowing continued avoidance of the host immune response³³⁸. In this situation there are multiple allelic forms of an antigen in a population at the same time. When the host generates and antibody response against a high frequency form this selects against isolates expressing this form reducing their frequency in the population. This simultaneously allows expansion of isolates expressing different allelic forms of the antigen. When this form reaches a high frequency it will again be selected against allowing expansion of another allelic form, the cycle can continue ad nauseam. Limited evidence of balancing selection at the epitope-level in this population identified by significantly positive values of Tajima's D was not supported by significantly negative values of Fay and Wu's H. Ct isolates are able to reinfect within households and communities with immunity only partial and being slow to develop. This means Ct is evading the host immune response, but not by balancing selection in a few immunodominant antigens. Therefore Ct must employ a currently unknown strategy for immune evasion.

8.2. Antibody responses and susceptibility to Ct infection

In this study serum from Gambian children with longitudinal evidence of infection and Gambian adults with trichomatous scarring were screened against microarrays of 908¹⁶⁸ and 894²⁰⁵ proteins of serovar D Ct to identify correlates of immunity in trachoma-endemic populations. In children diverse antibody responses, characterised by responses against a high number of antigens with no clear immunodominant targets, were associated with susceptibility to frequent and long duration ocular Ct infections. These non-immune children also had heightened responses against a panel of 42 antigens which associated with susceptibility to infection. Children with partial immunity to Ct infection had more focussed antibody responses, within these more focussed responses 5 potentially protective antibody responses were identified.

In adults with or without trichomatous scarring, antibody responses were more focussed than observed in children. This was highlighted by the targeting of antibody

responses against EB outer membrane proteins and proteins secreted into the host cytosol. These changes in antibody profile with age suggested long-term repeated exposure to Ct infection was driving out diversity and contributing to the development of only partial immunity. Heightened responses against 8 antigens were associated with scarring trachoma in this population, responses against just 1 antigen were associated with a lack of scarring.

These findings showed that antibody responses in trachoma-endemic populations change with exposure and are correlated with differential development of immunity to Ct infection. Despite the observation that antibody profiles are indicative of immunity, there was limited evidence of protective antibodies in children or adults. Notably, responses against MOMP and PmpD, which have been shown *in vitro*, to induce antibodies able to neutralise Ct infectivity^{132, 134}, were absent or infrequent. The lack of protective antibodies in a setting where antibody responses associated with the differential acquisition of immunity to Ct infection, combined with the absence of balancing selection, suggests that Ct employs a different strategy for immune evasion.

The results suggest two potential routes of immune evasion which would allow Ct reinfection while continuing to slow the development of immunity. These are the blocking hypothesis and the decoy hypothesis.

8.2.1. The blocking hypothesis

The blocking hypothesis is centred around the finding that of the 42 heightened responses which associated with susceptibility to infection in children, 3 were EB outer membrane proteins. The hypothesis is that antibody responses targeted against these surface proteins block the binding of protective and potentially neutralising antibodies (Figure 8.2).

Blocking antibodies have been found previously in Ct and others infections. A study of the pathogenic fungus *Candida albicans* identified differences in protection afforded by antibodies against cell-surface mannoproteins compared with antibodies targeted against the underlying β -glucans in mouse models²⁹⁶. Exposure of the β -glucans generated protective antibody responses which could be blocked by antibody responses targeted to the surface mannoproteins. Crane *et al* described a similar

phenomenon in Ct infection. Initially they showed that PmpD antiserum had pan-neutralising ability for Ct infectivity in ocular, urogenital and LGV serovars¹³⁴. This neutralising ability was severely diminished by pre-incubation of Ct EBs with antibodies specific to the immunodominant antigens MOMP or LPS, suggesting these antibodies could block the protective effects of PmpD antibodies. If anti-PmpD antibodies were incubated with Ct EBs prior to addition of anti-MOMP or anti-LPS antibodies, PmpD antibodies retained their neutralising ability. This demonstrated the presence of blocking antibodies in Ct and highlighted the importance of the interplay between antibodies, as the order in which the antibodies contacted the EBs significantly impacted their function.

The three susceptibility associated antigens were CT017 (Ctad1), CT541 (MIP) and CT579. The blocking hypothesis is that high levels of antibodies generated against these antigens in susceptible individuals bind the surface of Ct EBs, therefore blocking the binding and neutralisation of antibodies induced against PmpD or other unidentified targets of protective antibodies. Blocking of protective antibodies would increase susceptibility to long duration Ct infections and subsequent reinfection. In support of this, both MIP and CT579 have previously been identified as immunodominant antigens^{297, 298}. Neutralising antibodies targeting MIP have been demonstrated *in vitro*, it is possible that antibody responses against this protein are similar to those seen with MOMP. Antibodies targeting certain MOMP epitopes have some serovar-specific neutralising ability, but can block antibodies which afford greater pan-specific protection.

Another possibility within this hypothesis is that antibodies against some EB surface proteins could enhance Ct infectivity, which would promote longer survival in individuals. Ctad1 is known to be involved in EB attachment and induction of host-cell signals required for invasion²⁹⁹. It is plausible that antibodies binding Ctad1 could target EBs to host cells or directly induce uptake through cross-linking of Fc receptors.

One of the 8 antigens associated with trachomatous scarring in adults was CT314, proteomic analysis has identified this protein in the outer membrane complex of Ct EBs³²⁹. As a surface protein CT314 could induce blocking antibodies or antibodies which enhance infectivity. This would prolong Ct survival in individuals, driving persistent conjunctival inflammation which is the strongest risk factors consistently identified in the progression of trachomatous scarring.

In the blocking hypothesis, antibodies generated against these surface proteins provide a survival advantage for Ct. Therefore, mutations that change these surface proteins should be under purifying selection, because it is of benefit to Ct for the host immune response to recognise these antigens. Three of four outer membrane proteins with strong evidence of selection were under purifying selection. These were PmpC, PmpH and MOMP. The MOMP region with the clearest signal of purifying selection was found within variable domain 1 (VD1). An epitope in this region has previously been shown to induce neutralising antibodies against ocular Ct serovars^{132, 364}. This strongly suggests that *in vivo* if antibodies against this region are being generated they are not helping clear Ct infection, purifying selection in VD1 supports the blocking hypothesis. There was no clear evidence of selection in the 3 susceptibility associated surface antigens. Scarring-associated CT314 had evidence of purifying selection, which supports this antigen as a target of blocking antibodies.

In this hypothesis surface antigens which induce protective and neutralising antibodies would benefit the host, therefore mutations in the encoding genes should be under positive selection to evade immune recognition. There was no clear evidence of selection in PmpD. The only other outer membrane protein with evidence of selection was CT396 (HSP70), this gene was under positive selection. HSP70 is immunogenic in urogenital infection and unlike HSP60 responses does not seem to associate with tubal pathology^{170, 410}. It is possible HSP70 antibody responses can be protective in trachoma-endemic populations, although it is only a minor component of the EB surface⁴¹¹.

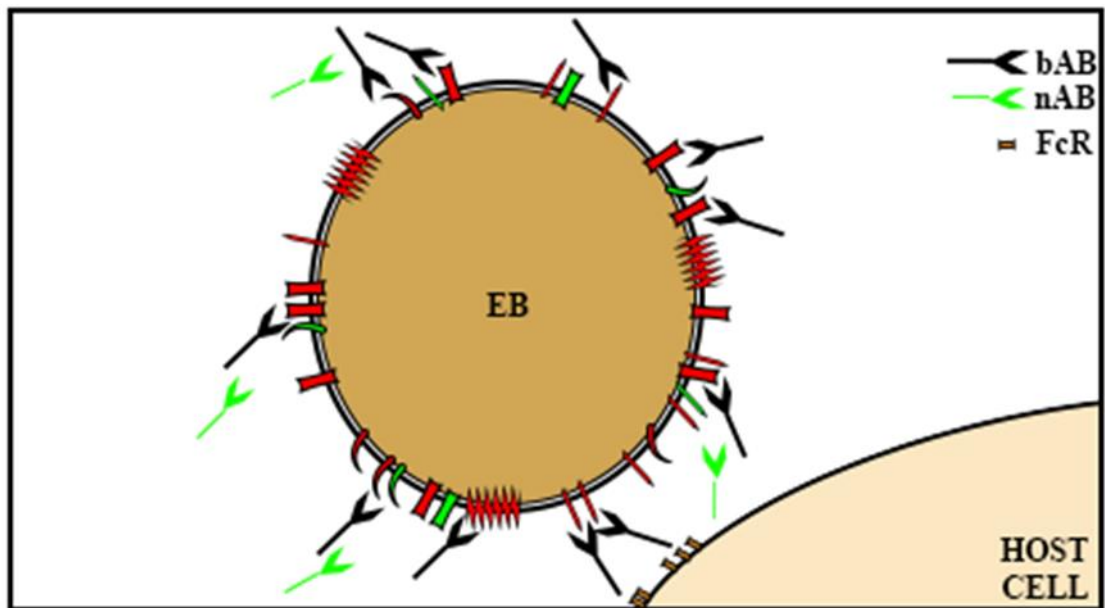


Figure 8.2. Blocking and enhancing antibodies.

Blocking antibodies (bAB) can physically block the binding of neutralising antibodies (nAB). These antibodies may also enhance infectivity of Ct through Fc receptors (FcR).

8.2.2. The decoy hypothesis

The majority of the 42 susceptibility associated and 7/8 of the scarring-associated antigens were not localised to the EB surface, therefore their role in Ct infection and disease cannot be explained by the blocking hypothesis. These responses show that a high number of Ct antigens are accessible by the host immune system, likely through release of Ct proteins as a result of cell lysis. Antibody responses against the vast majority are not protective. The decoy hypothesis is that heightened responses to a diverse panel of irrelevant antigens divert the humoral immune response, or at least a significant portion of it, away from potentially protective antigens. The frequent recognition of these non-protective antigens was reflected in their associations with susceptibility and scarring, suggesting these are not random but are deliberately presented by Ct to the host. Ct presents a diverse panel of antigens to the host to act as decoys, limiting development of protective antibody responses and diluting the effect of those that do develop (Figure 8.3).

This decoy evasion tactic relies on broad antibody responses to limit the development of protective antibodies, therefore individual responses would not be

expected to predict susceptibility to infection and downstream sequelae. This was supported by data from screening of sera from children and adults. Several antigens were identified where antibody responses were associated with susceptibility to infection or trachomatous scarring, however a combination of multiple responses was required to explain the observed differences in immunity.

Part of how this diverse antibody response may limit production of protective antibodies is related to the limitation of resources for B-cell proliferation and downstream impact on affinity maturation of anti-Ct antibodies. B-cell proliferation and production of antibodies is resource-limited and dependent on T-cell help. In individuals susceptible to infection where decoy antigens are stimulating broad antibody responses, resources are being 'wasted' on non-protective responses. This would limit the initial proliferation of B-cells stimulated by protective antigens. This is equally true for reactivation of plasma cells upon reinfection. Plasma cells have to compete for survival niches, the ability of these cells to be long-lasting is dependent on the initial levels of antigen-specific plasma cells and how often they encounter and are reactivated by the relevant antigen²⁹². If the initial proliferation of plasma cells is limited this would affect maintenance of these cells in the future, showing how initially diverse antibody responses could impact responses to reinfection and the development of immunity.

Affinity maturation is the process by which antibody affinity to its specific inducing antigen increases. This involves somatic hypermutation of variable regions within surface-bound antibodies on B-cells, which are then selected for based on their new affinity for the inducing antigen which are trapped and presented within lymphoid follicles. The slow development of immunity in Ct suggests considerable re-exposure and reactivation of antigen-specific cells is required. Affinity maturation of protective antibodies may well be important in this process, as seen with neutralising antibodies in HIV²⁹³⁻²⁹⁵ and also suggested by non-human primate models of trachoma¹³⁶. The decoy hypothesis is based upon the ready availability of high numbers of antigens and extensive production of antibodies. This abundance of antigens would significantly impact the threshold for affinity-based selection of B-cells in follicles during affinity maturation, weakening the power of selection to remove those with low affinity. Therefore, the diverse antibody response induced in the decoy hypothesis would also be partially self-sustaining through expansion of a large pool of non-protective B-cells, rather than affinity maturation of selective high-affinity protective responses.

Validation of the decoy hypothesis through evidence of natural selection is more challenging than for the blocking hypothesis. The diverse antibody profile would provide an advantage for the bacteria, however the number of antigens targeted and the inherently random nature of them implies no one gene would show strong evidence of selection. If selection was acting upon antigens targeted in this way, it would be expected to be purifying as recognition by the host immune system is beneficial for the bacteria.

In support of the decoy hypothesis 37/42 of the susceptibility associated antigens and 7/8 of the scarring associated antigens had no strong evidence of selection. The only scarring associated antigen under selection was CT314, as described previously this was under purifying selection. Two of the susceptibility associated antigens, CT545 and CT806, were also under purifying selection. CT228, CT694 and CT695 were the only susceptibility associated antigens with evidence of positive selection. In CT695 the regions under selection did not contain predicted epitopes. CT228 and CT695 had some overlap with regions under selection and predicted epitopes. As described previously they are both immunogenic but also functionally important for Ct exit and entry respectively. One limitation with detecting signatures of selection in genes involved in Ct survival is that determining the driving force of selection is difficult. For these two proteins it cannot be determined whether immune recognition or changes in Ct infectivity and transmission are the cause of selection. Overall, the antigens associated with sustained Ct infection and trachomatous scarring showed limited evidence of selection, supporting their group function as part of a panel of decoy antigens.

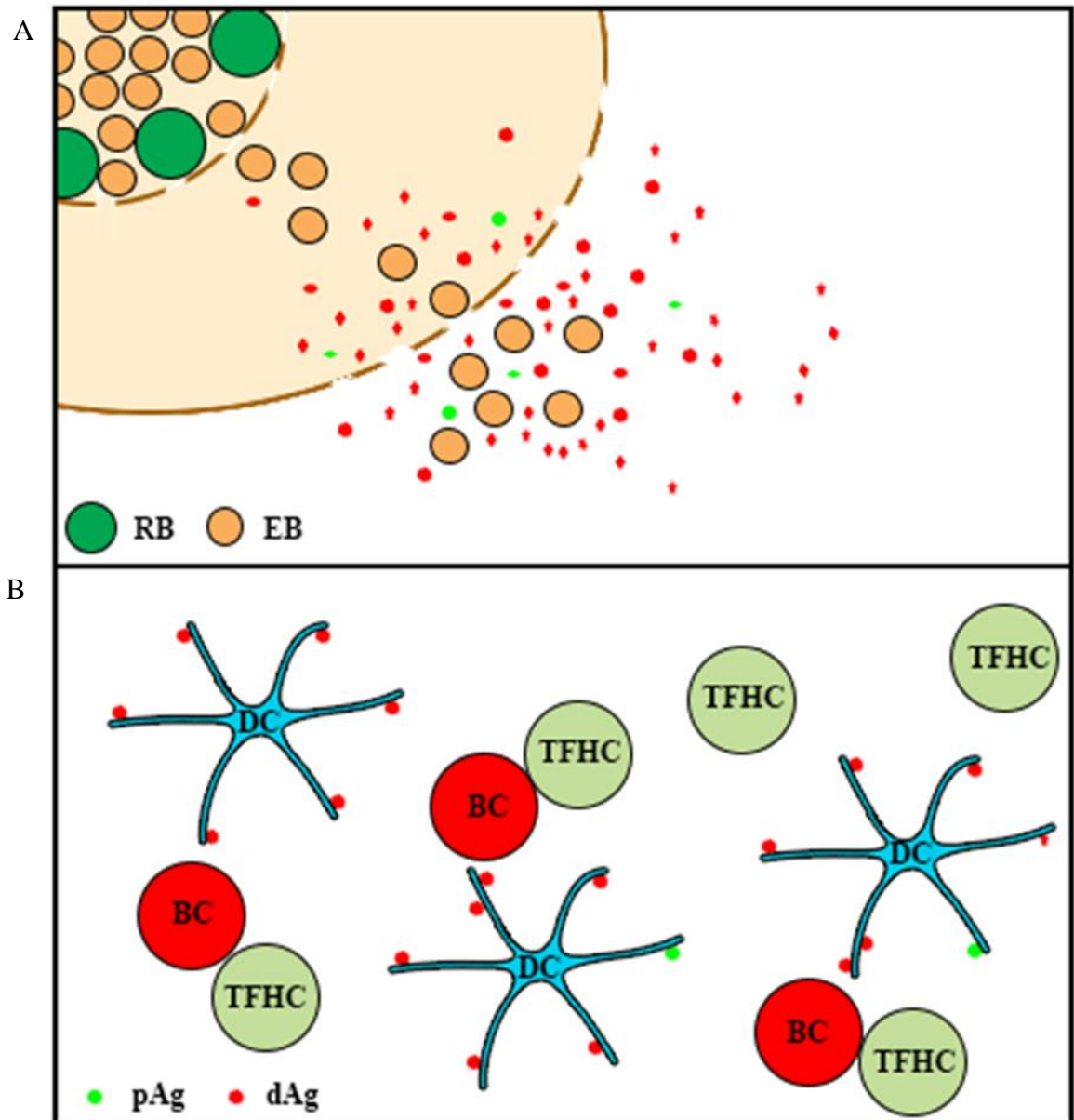


Figure 8.3. Decoy antigens divert antibody responses from protective antigens.

Ct presents a large pool of decoy antigens (dAg) to the immune system to divert antibody responses away from potentially protective antigens (pAg) (A). An excess of Ct dAg presented by follicular dendritic cells (DCs) and the ready availability of T-cell help (TFHC) in follicles induced by Ct infection, weakens the selectivity in affinity maturation and clonal expansion of B-cells (BC) (B).

8.2.3. Limitations of the natural immune response to ocular Ct infection

From studying children with differential evidence of immunity to Ct infection and adults with and without trachomatous scarring, there was limited detection of antibody responses associated with protection. The decoy hypothesis likely plays a role in this, by diverting host humoral immune responses from protective antigens. However even after grouping individuals based on this diverse antibody profile, children with more focussed responses still had limited evidence of protective antibodies. The five antigens putatively associated with protection from infection were not surface-exposed or localised to facilitate interaction with the host, they had also not been previously identified as immunogenic which casts doubts on their protective capacity. In adults who had no evidence of scarring there was still only one antigen, CT442, which was associated with a lack of scarring.

Individually protective and neutralising antibody targets have been found *in vitro* and through non-human primate models, little evidence from studies in humans from trachoma-endemic communities have corroborated these findings *in vivo*. This may in part be due to limitations of the micro-array approach for identifying candidate antigens. Using MOMP and PmpD as examples of antigens capable of inducing Ct infectivity neutralising antibodies, little reactivity was found against MOMP in particular in both studies. They are large, multimeric proteins which localise to the EB outer membrane *in vivo*^{306, 307}, PmpD also appears to undergo post-translational modifications resulting in expression of a smaller soluble form. It is unclear how the *in vitro* expression of these proteins on the micro-arrays resembles their natural confirmation, which may impact the exposure of certain epitopes.

While MOMP is clearly immunodominant, there has been limited evidence from human studies that protective antibody responses are stimulated *in vivo*. A longitudinal study from The Gambia approximately 10 years prior to the study described in chapter 4 found that heightened anti-MOMP IgG levels in tears were associated with acquisition of Ct infection²⁸⁰. This raises the question of how often antibodies against the protective and/or neutralising epitopes of MOMP and PmpD are stimulated in natural infection. Combined with the described impact of diverse antibody responses on immunity to Ct infection, an absence of these antibodies in natural infection within trachoma-endemic communities would help explain the slow and partial development of immunity. This poor development of natural immunity is supported by the observation that when

observed longitudinally, every 2 weeks for 13 visits, a quarter of adults in a Gambian trachoma-endemic community had one or more episodes of infection and these were associated with an increased proportion of intense inflammatory episodes compared with children aged under 15 years⁸. Therefore, while partial immunity to Ct may reduce the frequency and duration infections, it does not ablate them and does not effectively control resulting inflammation. If immunity led to the development of protective and neutralising antibodies, rather than the improved focussing and reduced diversity of antibody responses as proposed here, stronger evidence of protection from Ct infection would be expected.

8.3. CT442, Rab proteins and intracellular trafficking

Previous studies of the Inc CT442 have focussed on its role in host immune responses to Ct. In mouse models a CT442-derived peptide can stimulate cytolytic activity from CD8⁺ T-cells¹⁹⁵. After immunisation of mice with a Vaccinia virus vector expressing this peptide, CT442-specific CD8⁺ T-cells were stimulated and upon reactivation by Ct could provide a level of protection equivalent to immunisation with Ct-EBs³⁷⁸.

Antibody responses against CT442 have also been demonstrated in women with current or a previous Ct infection^{168, 169}. A non-human primate model of trachoma in which animals were partially immune to Ct re-challenge found reactivated antibody responses to be important, this included antibodies against a 15 kDa protein which was unidentified but could potentially be CT442. In contrast, CT442 functional characterisation has been minimal. It is known to be expressed mid-late during the Ct developmental cycle^{212, 213} and localised to the inclusion membrane⁸³, but no known functions have been described.

CT442 was highlighted here through serological studies and population genetics as being an important Ct protein in immunity and Ct fitness in trachoma-endemic populations. In The Gambia, heightened CT442 antibody levels were associated with an absence of scarring in adults. In Tanzania CT442 antibody responses were heightened in children whose conjunctival scarring had not progressed over a period of 4 years. Within a population of ocular Ct isolates collected on the Bijagos Islands, Guinea-Bissau²⁴⁰, CT442 was shown to contain regions under both positive and purifying selection.

8.3.1. CT442 is evolving under non-neutral selection

In the Ct population studied here CT442 is evolving under non-neutral selection, the forces driving this selection are difficult to identify. There are two regions within the gene under strong forces of selection. Positive selection is acting within the first predicted transmembrane domain. Purifying selection is acting within a C-terminal region predicted to be exposed to the host cytosol during Ct infection.

An absence of positive selection in an exposed region of CT442 was surprising, given the association of antibodies with a lack of scarring in adults and reduced scarring progression in children. If antibody responses against CT442 can limit scarring progression the most plausible method for this would be through antibody recognition of CT442 on the surface of extrusions. If CT442 was exposed on extrusions, antibodies could target these for destruction. This would reduce Ct transmission and therefore duration of infection, presumably helping to minimise damaging conjunctival inflammation. If this was the case mutations that help evade antibody recognition would be favoured, supporting positive selection of new alleles. Heightened antibody responses against CT442 could also indicate upregulated expression or increased availability of this antigen to the immune system generally, as expression of Incs is known to vary between serovars of Ct³³² and therefore may vary within ocular strains. It is plausible that heightened CT442 antibody responses in adults with scarring are indicative of increased accessibility to this antigen promoting aforementioned CD8⁺ T-cell responses. In support of this, CD8⁺ T-cells were recently identified as important in immunity against Ct in non-human primates previously vaccinated with a plasmid-deficient live-attenuated Ct strain²⁰⁰.

In relation to this hypothesis, the region of positive selection in CT442 could be related to the CD8⁺ T-cell inducing peptide found in mice. Ability to induce this response in humans has not been proven, however the peptide falls within the region under positive selection within this Ct population. Incs are believed to be secreted from the inclusion into the host cytosol prior to insertion into the inclusion membrane. This could provide an opportunity for CT442 to interact with the antigen presentation pathway for MHC-Class 1. If this CT442 peptide can induce cytolytic responses it is plausible that mutations in this peptide interfere with recognition, therefore providing a selective advantage to isolates with the new CT442 alleles driving these alleles to fixation as in positive selection.

The region of purifying selection did not overlap with predicted B-cell epitopes and was found in the putatively host-cytosol exposed C-terminal region. If immune responses are not driving this selection, this region may be important for the intracellular function of CT442.

8.3.2. Potential interactions of CT442

Co-immunoprecipitation of CT442-GFP from Ct-L2 infected cells identified a number of proteins and pathways that potentially interact with CT442. Intracellular trafficking was the strongest supported pathway and within this Rab proteins were the most abundant class of proteins.

Several Rab proteins have been shown to localised peripheral to the inclusion in Ct-infected cells, recent proteomic analysis of inclusions isolated from mammalian cells further supported the involvement of Rab proteins in Ct intracellular trafficking and interactions with host organelles. Currently few specific interactions are known. Rab11 interacts with IncG⁹⁰, which is thought to important for Ct-resistance to apoptosis⁹⁴. Rab11 together with Rab6 is also required for Ct-induced Golgi fragmentation and related nutrient acquisition⁹¹. Rab14 has been implicated in the delivery of sphingolipids to developing inclusions³⁹⁴.

Rab7 had one of the strongest scores from mass spectrometry analysis of proteins co-immunoprecipitated by CT442-GFP. Rab7 is traditionally associated with late endosomes and fusion with lysosomes³⁹⁵, it is now known to function in retrograde trafficking between endosomes and the Golgi³⁹⁶, recruitment of the proteasome and nutrient transport through lipid trafficking³⁹⁷. Rab7 is also involved in phagosomal maturation, different pathogens have developed novel ways to manipulate recruitment and expression of Rab7 to avoid lysosomal fusion and destruction³⁹⁷. If Ct infects macrophages Rab7 and LAMP1, another late endosomal marker, rapidly associate with the bacteria and suppress its growth³⁹⁸. It is possible CT442 functions in the exclusion of Rab7 from the inclusion during intracellular trafficking, although its mid-late cycle expression implies its functions later in the developmental cycle.

A recent study identified CT442 in the proteome of Ct-recruited lipid droplets. Lipid droplets are host organelles primarily involved in lipid storage, they appear to be

recruited to the inclusion and actively translocated across the inclusion membrane during Ct infection⁸⁶. Enrichment of lipid metabolism in these organelles and recruitment to the inclusion periphery was essential for Ct replication⁹³. Rab proteins Rab7, 11 and 31 were also identified in the Ct-lipid droplet proteome. All of these Rab proteins were co-immunoprecipitated by CT442-GFP. Lipid droplet recruitment is first detectable around 18 hours post infection, this matches the earliest expression stage of CT442. These three Rab proteins and another potential CT442-interactor Rab13 have all been implicated in endosomal trafficking to and from the trans-Golgi network involving recycling endosomes^{90, 399}. It is possible CT442 engages this pathway and/or lipid droplets through Rab proteins for nutrient acquisition and therefore Ct intracellular development.

8.4. Implications

8.4.1. Implications for a *Chlamydia trachomatis* vaccine

These studies identified focussed antibody profiles as associated with protection from Ct infection. Conversely individuals who developed a diverse antibody response against primarily non-protective antigens were susceptible to repeated and sustained Ct infections. Antibody responses became more focussed with age and the development of partial immunity. However, the non-surface localisation and lack of evidence of selection in antigens associated with scarring in adults suggests early development of ineffective antibody responses may be a driving factor in the repeated infections and subsequent persistent inflammation that drive trachomatous conjunctival scarring.

The limited observation of protective antibodies, notably targeting immunodominant antigens such as MOMP and PmpD which can induce neutralising antibodies *in vitro*, raised questions about the partial immunity to Ct which develops in trachoma-endemic communities. Natural immunity to ocular Ct infection is slow to develop and never fully protects from infection or conjunctival inflammation. Children who had partial immunity or were at least better protected than other children in this study had more focussed global responses, however they had little evidence of protective antibodies that would be expected from *in vitro* data. This discrepancy

between natural responses to Ct infection and the immune responses believed to help clear Ct has important implications for vaccine design.

Current developments towards a Ct vaccine can be broadly grouped into those which utilise a modified or inactivated form of Ct-EBs and those which use whole or sub-units of immunogenic proteins. Studies in mice and non-human primates have shown partial protection from re-challenge with Ct using UV-inactivated Ct and plasmid-deficient Ct respectively^{136, 189, 200, 201}. Both methods have identified T-cells as important, the latter also showed reactivation of antibody response against a panel of Ct antigens to be important in protection. Whole protein or sub-unit vaccines almost exclusively utilise MOMP^{191, 412} and Pmp's specifically PmpC¹⁸⁷ and PmpD¹⁹². These studies have found induction of both cellular and humoral immune responses to be important in clearance. They have also found the development of antibodies against these antigens can neutralise Ct infectivity.

The results presented here strongly support the use of a sub-unit based vaccine for future Ct development. Natural exposure to Ct-EBs does not generate a protective response similar to that seen in mouse models. A vaccine based on inactivated or modified EBs would likely induce variable outcomes in different individuals, as seen with natural infection. Additionally, as highlighted and previously shown, antibody responses against EB-surface antigens are not universally protective. Antibodies targeting outer membrane proteins were associated with susceptibility to infection children and the presence of scarring in adults. The encoding genes either had evidence of purifying selection or none at all, indicating at best that responses against don't benefit the host and at worst that they facilitate enhanced Ct infectivity and survival.

The limited evidence of individually protective antibodies in these studies means the potential efficacy of sub-unit vaccines can only be speculated upon. They would promote a highly focussed and targeted immune response. Evidence presented here suggests that the diversity of antibody responses is negatively correlated with protection from infection. Therefore a vaccine which induces responses against one or a few protective antigens would likely have a greater chance of inducing long-lived immunity.

8.4.2. Implications for Ct genomics

The use of population genetics to identify signatures of selection within 126 ocular Ct isolates from the Bijagos Islands, Guinea-Bissau was the first of its kind in *Chlamydia*. Previous studies have utilised from 3 to 59 Ct genomes to look for positive selection in Ct as a whole and within the biovars^{114, 115, 355, 356}, however discovery of genes under selection within a restricted, currently circulating Ct population is a novel study and result. Genes under positive selection in this study suggests they are providing a selective advantage for Ct survival and transmission within this population.

In an era when whole-genome sequencing is becoming increasingly affordable and accessible, this study has important implications for the study of Ct transmission dynamics and evolution. Studies historically sequenced *ompA* to study the transmission of Ct in communities¹⁰⁷⁻¹⁰⁹. The evidence for purifying selection found here and previous evidence showing frequent recombination involving this locus suggests that changes in *ompA* do not reflect the ongoing diversification in Ct populations. This study suggests it may be possible to track transmission through panels of genes known to be evolving under natural selection.

Population genetics may become a tool for identifying candidate virulence factors or immunologically important targets. Genes under selection in this population included outer membrane proteins, secreted effectors and inclusion membrane proteins which are known to be important in Ct invasion, intracellular development and transmission. Many of those under selection have no known function but are localised to facilitate interactions with the host. Evidence of selection in these targets highlights these genes as important factors in Ct fitness, suggesting they should be targeted for functional studies. Similarly evidence of selection can highlight and validate immune targets. Combined with micro-array identification of proteins which stimulate antibody responses, evidence of selection can filter out antigens where responses against them provide a demonstrable advantage or disadvantage for Ct.

8.5. Limitations

8.5.1. Microarray limitations

Chapters 4 and 5 highlighted the importance of validation of targets from high-throughput, serological microarrays. Antibody responses from independent testing, by ELISA, of target antigens showed mixed correlation with microarray results. Furthermore, testing of serum from independent but clinically related individuals did not consistently support the association of specific antibody responses with clinical outcomes. These results suggest some of the differentially recognised antigens identified were likely false-positive artefacts from screening of small numbers of sera on the array. These results demonstrated the problems of small sample sizes, a common issue due to the relatively high cost of arraying samples versus single-antigen ELISA. They also underline the need for clearly defined clinical phenotypes in human studies of humoral immunity to improve the reliability of statistical associations and reduce the influence of outliers.

A further limitation of the microarrays was the failure to identify responses against known Ct immunodominant surface antigens. MOMP and the majority of polymorphic membrane proteins (pmp) were minimally immunogenic, since they did not pass initial filtering of recognised antigens. This was likely due to poor expression of these proteins, as they are large, structurally complex proteins. PmpC (CT414) and PmpD (CT812) were the only Pmp's frequently recognised and both were expressed as fragments, not full-length proteins. This suggests large, multimeric proteins such as these should be expressed as fragments or peptides alongside full-length proteins in microarrays. Single/multi-antigen ELISA or immunoblotting against Ct EBs should also be considered for known, immunodominant antigens whose microarray expression is predicted to be unable to capture their target antibodies.

8.5.2. Influence of urogenital chlamydial infection

Urogenital Ct infection, and possible maternal antibodies, should not be a factor in chapter 4, since the participants were mostly aged between 6 and 12 years. In contrast,

present and more likely past urogenital infections may have had some impact on the antibody responses determined in individuals studied in chapter 5, who were aged between 30 and 80 years. High levels of sequence conservation for the majority of Ct proteins, demonstrated by approximately 99% sequence identity of currently sequenced isolates, may lead to cross-reactivity of antibodies generated against ocular and urogenital infections.

There are reasons to believe the potential effect of antibodies generated against urogenital Ct infection were not a significant confounder in this work. Historically minimal evidence of urogenital Ct infection in The Gambia is the primary support for this belief, although there is not much literature on the subject. Studies from The Gambia⁴¹³ and a rural region of Senegal⁴¹⁴ found urogenital Ct infection levels to be < 2%. Similarly, a broader study of low-middle income countries estimated urogenital Ct prevalence at 2.7%⁴¹⁵. Urogenital infection at this prevalence would be unlikely to have significantly influenced the results here. Furthermore, survey designs and adjusting for age and gender should have reduced the likelihood for systematic differences in urogenital exposure to exist between the scarring cases and healthy controls studied in chapter 5.

Since there is not equivalently well-described history of urogenital infection for The Gambia, this should be considered further in the future. It may be possible, by evaluating antibody responses against serovar specific antigens, to adjust for exposure to urogenital Ct. Previously the MIF test has been used for this purpose, however there is known to be considerable cross-reactivity between ocular and urogenital serovars and with other chlamydial species⁴¹⁶. An assay utilising shorter peptides may be able to provide greater resolution, a collaborator is currently investigating this²⁰⁴.

8.5.3. Interplay of humoral and cellular immunity

The weight of literature in Ct infections clearly identifies the importance of the cell-mediated immune responses in the resolution of infection. CD4⁺ and CD8⁺ T-cell anti-Ct responses have been identified as important in infection resolution^{125, 126, 129, 200, 417} and differential progression of scarring pathology^{126, 418} by longitudinal studies of ocular

and urogenital Ct. In contrast, the role of humoral responses in human cohort studies is less clear. Levels of local mucosal and serum anti-Ct antibodies are higher in children who experience increased rates of ocular reinfection^{279, 280} and more severe trachomatous inflammation^{271, 279, 419}.

However, there is considerable evidence from animal models suggesting antibody responses are also necessary for long-term protection and immunity from infection. Olsen *et al* demonstrated the development of Ct-infection neutralising antibodies, dependent on CD4⁺ T-cells, in a murine model following mucosal vaccination of a MOMP-based vaccine¹⁸⁸. This corroborated earlier murine studies which found B-cells and antibodies to be necessary for a complete cell-mediated immune response and associated long-term protection. In agreement with findings from our study, Teng *et al* showed Th1-favouring C57BL/6 mice, which are more resistant to chlamydial infection⁴²⁰⁻⁴²², had a more focussed antibody profile after *C. muridarum* infection⁴²³. Similarly in non-human primates, partial immunity to ocular infection was concurrent with development of a focussed antibody recall response¹³⁶.

The non-protective antibody responses identified cannot completely explain observed differences in immunity, as evidenced by the heterogeneity of antibody profiles within the protected and susceptible groups in chapter 4. Studies of cell-mediated responses in Ct infection have similarly not identified correlates of immunity which can completely explain observed variability of outcomes. Previous work on the cohort studied in chapter 4 found increased conjunctival expression of IFN γ , TNF α and IL-10²³⁵ and detectable MOMP-specific CD8⁺ T-cells⁴²⁴ associated with longer duration of infections. Together, the results reflect the complex relationship between cell-mediated immune responses, which are necessary for initial clearance of Ct and those required for long-term protection. The results presented here support the importance of antibody responses in human immunity to Ct infection but as expected they cannot fully

account for the complete immunological picture in the absence of combined measures of cell-mediated immune responses.

8.6. Future work

8.6.1. Serological validation

Future work should aim to validate the blocking and decoy hypotheses proposed. This work should be focussed on longitudinal studies, the temporal information they provide is necessary in understanding how antibody responses and immunity develop in relation to Ct exposure and trachomatous disease. The blocking hypothesis should be tested using a selection of EB surface antigens, including scarring-associated CT314. *In vitro* studies similar to those which identified blocking antibodies against PmpD should be paired with serological identification of responses against these antigens in trachoma-endemic populations.

Testing of the decoy hypothesis likely requires further large-scale micro-array screening of children from trachoma-endemic communities. An earlier longitudinal study from The Gambia could be used initially²⁸⁰, if global antibody profiles maintain their association with differential evidence of immunity then new populations would need to be tested. This previous longitudinal study also showed that tear and serum antibodies may not be correlated in trachoma. Tears were collected and are available from every 2 week visit from the study described in chapter 4. Micro-array screening of tears is not practical because of low antibody levels, initial studies focussing on a few specific targets should be performed.

While identification of individually protective responses was limited, two antigens associated with susceptibility to infection (CT694 and CT228) and the only antigen associated with protection from progressive scarring (CT442) had evidence of selection suggesting they should be investigated further. CT228 and CT442 also highlight the need for investigation of how Incs are targeted by the humoral immune response. Several Incs appear to be immunogenic in urogenital and/or ocular Ct infection³⁷⁰, yet no work has been undertaken to understand how they are exposed to the extracellular space to be targeted by antibodies.

8.6.2. Independent populations for Ct genomics

Future population genomics studies should be undertaken in populations of ocular Ct isolates to determine whether genes under selection are shared. A limitation of the present study is that understanding the driving force of selection in these genes is difficult and necessarily involves a level of speculation based on the available information for each gene and the proteins they encode. Identification of genes under selection in other ocular Ct populations could help decipher what is driving selection. Genes under selection driven by impact on Ct survival should be common to ocular Ct populations, those under immune response-driven selection may vary depending on the levels of transmission and endemicity.

The lack of clear evidence for balancing selection in this population should also be investigated further as it has important implications for understanding Ct evasion of host immune responses. As well as whole-genome scans, antigens known to be immunodominant in Ct should also be investigated by targeted sequencing, as has been done previously for *ompA*.

8.6.3. Understanding the role of CT442 in the Ct developmental cycle

The first step in further characterisation of CT442 should utilise the anti-CT442 antibody to co-immunoprecipitate endogenous CT442 from mammalian cells to validate proteins identified using exogenously expressed CT442 through transfection of CT442-GFP into host cells. Anti-Rab7 antibody should be used to determine if endogenous CT442 also interacts with Rab7. Dependent on the outcome of this initial validation step, either Rab7 or interactions with lipid droplets should be pursued.

If the interaction is confirmed the localisation of Rab7 should be investigated in Ct infected mammalian cells. Rab7 interactions with other bacteria have been shown to vary considerably with different tissues and species³⁹⁷. Therefore, localisation should be investigated in conjunctival and genital epithelial cell lines with ocular and urogenital

serovars. The impact of expression of a Rab7 dominant negative mutant on Ct infectivity should also be investigated.

To examine interactions with lipid droplets, these organelles should be induced as described previously and examined in cells stained with anti-CT442 antibody to determine any co-localisation. This work and the proposed studies with Rab7 should also be performed using serovar C Ct strains which appear to express a truncated CT442. Interpretations from these studies would however be complicated because other genes contain truncating mutations in these strains, and there may be a redundancy of function in strains without a functional CT442.

8.7. Summary

In children from trachoma-endemic communities in The Gambia, the global profile of anti-*Chlamydia trachomatis* was correlated with development of immunity to Ct infection. Children with diverse antibody responses are more susceptible to recurrent and longer duration ocular Ct infections. Children who had more focussed antibody profiles were protected from infection. Breadth and diversity of antibody responses were driven by the deliberate presentation of large numbers of irrelevant, non-protective antigens to the humoral immune system by Ct to act as decoys, diverting antibody responses and associated resources away from potentially protective epitopes. This susceptibility-associated global profile was worsened by the induction of antibodies against non-protective EB surface antigens, which block the binding and neutralising ability of protective antibody responses. Evidence of purifying selection in EB surface antigens and an absence of selection in the majority of antigens associated with susceptibility to infection supported the decoy and blocking hypotheses as explanations for how antibody responses against Ct develop during natural ocular infection.

Continued re-exposure with age was characterised by a focussing of antibody responses in adults with and without trachomatous scarring. The limited identification of individual antibody responses which distinguished between adults with and without scarring suggested the differential development of antibody responses in early life, specifically through the recognition of antigenic decoys, was important in driving conjunctival scarring. In support of this a scarring-associated antigen (CT706) was associated with conjunctival scarring progression in children from a geographically

distinct trachoma-endemic population in northern Tanzania. CT442 which protected from scarring was associated with protection from progression in children in this study. There was no evidence of selection in the scarring-associated antigens, except a single EB-surface antigen (CT314) which was under purifying selection, further supporting the decoy and blocking hypotheses.

The discovery of genes under selection within a restricted, currently circulating Ct population was a novel finding. Evidence of selection validated well characterised Ct virulence factors, including TARP, CT223, and immunologically relevant antigens, including MOMP and CT694. This scan also identified likely virulence factors which are currently not characterised. Balancing selection was notably limited in this population, suggesting Ct employs different strategies for immune evasion including the use of decoy antigens and blocking antibodies. Several genes had evidence of both positive and purifying selection implying multiple driving forces of selection were acting, these included CT442 which was associated with protection from trachomatous and progressive trachomatous scarring.

CT442 was confirmed as an inclusion membrane protein with potential interactions focussed in intracellular trafficking and Rab proteins. Evidence of selection in CT442 did not appear to overlap with B-cell epitopes, meaning selection pressure related to impacts on Ct intracellular survival and transmission were most likely the driving force in evolution of CT442 in this population.

References

1. Mabey, D.C.W., A.W. Solomon, and A. Foster, *Trachoma*. The Lancet, 2003. **362**(9379): p. 223-229.
2. Grayston, J.T., et al., *Importance of Reinfection in the Pathogenesis of Trachoma*. Reviews of Infectious Diseases, 1985. **7**(6): p. 717-725.
3. Silverstein, A.M., *The immunologic modulation of infectious disease pathogenesis*. Investigative Ophthalmology 1973. **13**(8): p. 560-574.
4. Thyelfors, B., et al., *Global data on blindness*. Bulletin of the World Health Organisation, 1995. **73**(1): p. 115-121.
5. Resnikoff, S., et al., *Global data on visual impairment in the year 2002*. Bulletin of the World Health Organisation, 2004. **82**(11): p. 844-851.
6. Pascolini, D. and S.P. Mariotti, *Global estimates of visual impairment: 2010*. British Journal of Ophthalmology, 2012. **96**(5): p. 614-618.
7. Burton, M.J. and D. Mabey, *The Global Burden of Trachoma: A Review*. PLoS Neglected Tropical Diseases, 2009. **3**(10): p. e460.
8. Bailey, R.L., et al., *The duration of human ocular Chlamydia trachomatis infection is age dependent*. Epidemiology and Infection, 1999. **123**(3): p. 479-486.
9. Grassly, N.C., et al., *The natural history of trachoma infection and disease in a Gambian cohort with frequent follow-up*. PLoS Neglected Tropical Diseases, 2008. **2**(12): p. e341.
10. Polack, S., et al., *Mapping the global distribution of trachoma*. Bull World Health Organ, 2005. **83**(12): p. 913-9.
11. WHO, *Global WHO Alliance for the Elimination of Blinding Trachoma by 2020*. Weekly epidemiological record, 2012. **87**(17): p. 161-168.
12. Smith, J.L., et al., *The geographical distribution and burden of trachoma in Africa*. PLoS Neglected Tropical Diseases, 2013. **7**(8): p. e2359.
13. Smith, J.L., et al., *Mapping the global distribution of trachoma: why an updated atlas is needed*. PLoS Neglected Tropical Diseases, 2011. **5**(6): p. e973.
14. Emerson, P.M., et al., *The SAFE strategy for trachoma control: using operational research for policy, planning and implementation*. Bulletin of the World Health Organisation, 2006. **84**(8): p. 613-619.
15. WHO, *Global Initiative for the Elimination of Avoidable Blindness*. 2000, World Health Organisation: Geneva, Switzerland.
16. Solomon, A.W., et al., *Trachoma control: a guide for programme managers*. 2006, World Health Organisation: Geneva, Switzerland. p. 1-53.
17. Solomon, A.W., et al., *Mass Treatment with Single-Dose Azithromycin for Trachoma*. The New England Journal of Medicine, 2004. **351**(19): p. 1962-1971.
18. Solomon, A.W., et al., *Two Doses of Azithromycin to Eliminate Trachoma in a Tanzanian Community*. The New England Journal of Medicine, 2008. **358**(17): p. 1870-1871.
19. Emerson, P.M., et al., *Review of the evidence base for the 'F' and 'E' components of the SAFE strategy for trachoma control*. Trop Med Int Health, 2000. **5**(8): p. 515-27.
20. Harding-Esch, E.M., et al., *Risk factors for active trachoma in The Gambia*. Transactions of the Royal Society of Tropical Medicine and Hygiene, 2008. **102**(12): p. 1255-1262.
21. West, S.K., et al., *Epidemiology of ocular chlamydial infection in a trachoma-hyperendemic area*. J Infect Dis, 1991. **163**(4): p. 752-6.

22. Lee, D.C., et al., *Seasonal effects in the elimination of trachoma*. Am J Trop Med Hyg, 2005. **72**(4): p. 468-70.
23. Hsieh, Y.H., et al., *Risk factors for trachoma: 6-year follow-up of children aged 1 and 2 years*. Am J Epidemiol, 2000. **152**(3): p. 204-11.
24. Last, A.R., et al., *Risk Factors for Active Trachoma and Ocular Chlamydia trachomatis Infection in Treatment-Naive Trachoma-Hyperendemic Communities of the Bijagos Archipelago*. PLoS Neglected Tropical Diseases, 2014. **8**(6): p. e2900.
25. Emerson, P.M., et al., *Transmission ecology of the fly Musca sorbens, a putative vector of trachoma*. Trans R Soc Trop Med Hyg, 2000. **94**(1): p. 28-32.
26. Abdou, A., et al., *Prevalence and risk factors for trachoma and ocular Chlamydia trachomatis infection in Niger*. Br J Ophthalmol, 2007. **91**(1): p. 13-7.
27. Burton, M.J., et al., *Bacterial infection and trachoma in the gambia: a case control study*. Investigative Ophthalmology and Visual Science, 2007. **48**(10): p. 4440-4444.
28. Edwards, T., et al., *Risk factors for active trachoma and Chlamydia trachomatis infection in rural Ethiopia after mass treatment with azithromycin*. Trop Med Int Health, 2008. **13**(4): p. 556-65.
29. Harding-Esch, E.M., et al., *Trachoma Prevalence and Associated Risk Factors in The Gambia and Tanzania: Baseline Results of a Cluster Randomised Controlled Trial*. PloS Neglected Tropical Diseases, 2010. **4**(11): p. e861.
30. Everett, K.D., R.M. Bush, and A.A. Andersen, *Emended description of the order Chlamydiales, proposal of Parachlamydiaceae fam. nov. and Simkaniaceae fam. nov., each containing one monotypic genus, revised taxonomy of the family Chlamydiaceae, including a new genus and five new species, and standards for the identification of organisms*. Int J Syst Bacteriol, 1999. **49 Pt 2**: p. 415-40.
31. Stephens, R.S., et al., *Divergence without difference: phylogenetics and taxonomy of Chlamydia resolved*. FEMS Immunology and Medical Microbiology, 2009. **55**(2): p. 115-119.
32. Stephens, R.S., *Chlamydial evolution: a billion years and counting*, in *Chlamydial Infections: Proceedings of the Tenth International Symposium on Human Chlamydial Infections*, J. Schachter, et al., Editors. 2002, International Chlamydia Symposium, 2002. p. 3-12.
33. Taylor, H.R., *Doyne Lecture: trachoma, is it history?* Eye (Lond), 2009. **23**(11): p. 2007-22.
34. Nunes, A., et al., *Adaptive Evolution of the Chlamydia trachomatis Dominant Antigen Reveals Distinct Evolutionary Scenarios for B- and T-cell Epitopes: Worldwide Survey*. PLoS One, 2010. **5**(10): p. e13171.
35. Humphrys, M.S., et al., *Simultaneous transcriptional profiling of bacteria and their host cells*. PLoS One, 2013. **8**(12): p. e80597.
36. Stephens, R.S., *Chlamydial Genomics and Vaccine Antigen Discovery*. The Journal of Infectious Diseases, 2000. **181**(Supplementary 3): p. S521-S523.
37. Harris, S.R., et al., *Whole-genome analysis of diverse Chlamydia trachomatis strains identifies phylogenetic relationships masked by current clinical typing*. Nature Genetics, 2012. **44**(4): p. 413-419.
38. Andersson, P., et al., *Chlamydia trachomatis from Australian Aboriginal people with trachoma are polyphyletic composed of multiple distinctive lineages*. Nat Commun, 2016. **7**: p. 10688.
39. Collier, L.H. and J. Sowa, *Isolation of trachoma virus in embryonate eggs*. Lancet, 1958. **1**(7028): p. 993-6.

40. Darougar, S., et al., *Isolation of TRIC agent (Chlamydia) in irradiated McCoy cell culture from endemic trachoma in field studies in Iran. Comparison with other laboratory tests for detection of Chlamydia*. Br J Ophthalmol, 1971. **55**(9): p. 591-9.
41. Darougar, S., et al., *Isolation of Chlamydia trachomatis from different areas of conjunctiva in relation to intensity of hyperendemic trachoma in school children in Southern Tunisia*. Br J Ophthalmol, 1979. **63**(2): p. 110-2.
42. Tang, F.F., et al., *Studies on the etiology of trachoma with special reference to isolation of the virus in chick embryo*. Chin Med J, 1957. **75**(6): p. 429-47.
43. Collier, L.H., S. Duke-Elder, and B.R. Jones, *Experimental trachoma produced by cultured virus*. Br J Ophthalmol, 1958. **42**(12): p. 705-20.
44. Smith, C.H., *Accidental laboratory infection with trachoma*. Br J Ophthalmol, 1958. **42**(12): p. 721-22.
45. Collier, L.H., S. Duke-Elder, and B.R. Jones, *Experimental trachoma produced by cultured virus. Part II*. Br J Ophthalmol, 1960. **44**: p. 65-88.
46. Bastidas, R.J., et al., *Chlamydial intracellular survival strategies*. Cold Spring Harbour Perspectives in Medicine, 2013. **3**(5): p. a010256.
47. Abromaitis, S. and R.S. Stephens, *Attachment and entry of Chlamydia have distinct requirements for host protein disulfide isomerase*. PLoS Pathogens, 2009. **5**(4): p. e1000357.
48. Cocchiaro, J.L. and R.H. Valdivia, *New insights into Chlamydia intracellular survival mechanisms*. Cell Microbiol, 2009. **11**(11): p. 1571-8.
49. Mueller, K.E., G.V. Plano, and K.A. Fields, *New Frontiers in Type III Secretion Biology: the Chlamydia Perspective*. Infection and Immunity, 2014. **82**(1): p. 2-9.
50. Nans, A., H.R. Saibil, and R.D. Hayward, *Pathogen-host reorganisation during Chlamydia invasion revealed by cryo-electron tomography*. Cellular Microbiology, 2014.
51. Burkinshaw, B.J. and N.C. Strynadka, *Assembly and structure of the T3SS*. Biochimica et Biophysica Acta, 2014.
52. Elwell, C.A., et al., *RNA interference screen identifies Abl kinase and PDGFR signaling in Chlamydia trachomatis entry*. PLoS Pathog, 2008. **4**(3): p. e1000021.
53. Mehltitz, A., et al., *Tarp regulates early Chlamydia-induced host cell survival through interactions with the human adaptor protein SHC1*. The Journal of Cell Biology, 2010. **190**(1): p. 143-157.
54. Kun, D., et al., *Chlamydia inhibit host cell apoptosis by inducing Bag-1 via the MAPK/ERK survival pathway*. Apoptosis, 2013. **18**(9): p. 1083-1092.
55. Derrick, T., et al., *Conjunctival MicroRNA expression in inflammatory trachomatous scarring*. PLoS Neglected Tropical Diseases, 2013. **7**(3): p. e2117.
56. Chen, Y.S., et al., *The Chlamydia trachomatis Type III Secretion Chaperone Slc1 Engages Multiple Early Effectors, Including TepP, a Tyrosine-phosphorylated Protein Required for the Recruitment of CrkI-II to Nascent Inclusions and Innate Immune Signaling*. PLoS Pathogens, 2014. **10**(2): p. e1003954.
57. Scidmore, M.A., E.R. Fischer, and T. Hackstadt, *Restricted Fusion of Chlamydia trachomatis Vesicles with Endocytic Compartments during the Initial Stages of Infection*. Infection and Immunity, 2003. **71**(2): p. 973-984.
58. Kim, J.H., et al., *Endosulfatases SULF1 and SULF2 limit Chlamydia muridarum infection*. Cellular Microbiology, 2013. **15**(9): p. 1560-1571.

59. Kim, J.H., et al., *Chlamydia trachomatis Co-opts the FGF2 Signaling Pathway to Enhance Infection*. PLoS Pathogens, 2011. **7**(10): p. e1002285.
60. Mehrlitz, A. and T. Rudel, *Modulation of host signaling and cellular responses by Chlamydia*. Cell Communication and Signaling, 2013. **11**(90).
61. Hybiske, K. and R.S. Stephens, *Mechanisms of Chlamydia trachomatis entry into nonphagocytic cells*. Infection and Immunity, 2007. **75**(8): p. 3925-3934.
62. Chukwuemeka Ajonuma, L., et al., *CFTR is required for cellular entry and internalization of Chlamydia trachomatis*. Cell Biology International, 2010. **34**(6): p. 593-600.
63. Hower, S., K. Wolf, and K.A. Fields, *Evidence that CT694 is a novel Chlamydia trachomatis T3S substrate capable of functioning during invasion or early cycle development*. Molecular Microbiology, 2009. **72**(6): p. 1423-1437.
64. Sisko, J.L., et al., *Multifunctional analysis of Chlamydia-specific genes in a yeast expression system*. Molecular Microbiology, 2006. **60**(1): p. 51-66.
65. Wehrl, W., et al., *From the inside out-processing of the Chlamydial autotransporter PmpD and its role in bacterial adhesion and activation of human host cells*. Mol Microbiol, 2004. **51**(2): p. 319-34.
66. Kari, L., et al., *Chlamydia trachomatis Polymorphic Membrane Protein D is a Virulence Factor Involved in Early Host Cell Interactions*. Infection and Immunity, 2014.
67. Su, H., et al., *A recombinant Chlamydia trachomatis major outer membrane protein binds to heparan sulfate receptors on epithelial cells*. Proc Natl Acad Sci U S A, 1996. **93**(20): p. 11143-8.
68. Fadel, S. and A. Eley, *Chlamydia trachomatis OmcB protein is a surface-exposed glycosaminoglycan-dependent adhesin*. J Med Microbiol, 2007. **56**(Pt 1): p. 15-22.
69. Clausen, J.D., et al., *Chlamydia trachomatis utilizes the host cell microtubule network during early events of infection*. Molecular Microbiology, 1997. **25**(3): p. 441-449.
70. Zhu, H., et al., *Persistent and acute chlamydial infections induce different structural changes in the Golgi apparatus*. Int J Med Microbiol, 2014. **304**(5-6): p. 577-85.
71. Heuer, D., et al., *Chlamydia causes fragmentation of the Golgi compartment to ensure reproduction*. Nature, 2008. **457**(7230): p. 731-735.
72. Hybiske, K. and R.S. Stephens, *Mechanisms of host cell exit by the intracellular bacterium Chlamydia*. Proceedings of the National Academy of Sciences of the United States America, 2007. **104**(27): p. 11430-11435.
73. Scidmore, M.A., *Recent advances in Chlamydia subversion of host cytoskeletal and membrane trafficking pathways*. Microbes and Infection, 2011. **13**(6): p. 527-535.
74. Boncompain, G., et al., *The Intracellular Bacteria Chlamydia Hijack Peroxisomes and Utilize Their Enzymatic Capacity to Produce Bacteria-Specific Phospholipids*. PloS One, 2014. **9**(1): p. e86196.
75. Dumoux, M., et al., *Making connections: snapshots of chlamydial type III secretion systems in contact with host membranes*. Current Opinion in Microbiology, 2014. **23C**: p. 1-7.
76. Dumoux, M., et al., *Chlamydiae assemble a pathogen synapse to hijack the host endoplasmic reticulum*. Traffic, 2012. **13**(12): p. 1612-1627.
77. Betts-Hampikian, H.J. and K.A. Fields, *The Chlamydial Type III Secretion Mechanism: Revealing Cracks in a Tough Nut*. Frontiers in Microbiology, 2010. **1**(114).

78. Brinkworth, A.J., et al., *Chlamydia trachomatis* *Slc1* is a type III secretion chaperone that enhances the translocation of its invasion effector substrate TARP. *Molecular Microbiology*, 2011. **82**(1): p. 131-144.
79. Kleba, B. and R.S. Stephens, *Chlamydial effector proteins localized to the host cell cytoplasmic compartment*. *Infection and Immunity*, 2008. **76**(11): p. 4842-4850.
80. Huang, Z., et al., *Structural basis for activation and inhibition of the secreted chlamydia protease CPAF*. *Cell Host Microbe*, 2008. **4**(6): p. 529-542.
81. Chen, A.L., et al., *CPAF: a Chlamydial protease in search of an authentic substrate*. *PLoS Pathogens*, 2012. **8**(8): p. e1002842.
82. Pennini, M.E., et al., *Histone methylation by NUE, a novel nuclear effector of the intracellular pathogen Chlamydia trachomatis*. *PLoS Pathogens*, 2010. **6**(7): p. e1000995.
83. Bannantine, J.P., et al., *A secondary structure motif predictive of protein localization to the chlamydial inclusion membrane*. *Cellular Microbiology*, 2000. **2**(1): p. 35-47.
84. Carabeo, R.A., D.J. Mead, and T. Hackstadt, *Golgi-dependent transport of cholesterol to the Chlamydia trachomatis inclusion*. *Proc Natl Acad Sci U S A*, 2003. **100**(11): p. 6771-6.
85. Beatty, W.L., *Trafficking from CD63-positive late endocytic multivesicular bodies is essential for intracellular development of Chlamydia trachomatis*. *J Cell Sci*, 2006. **119**(Pt 2): p. 350-9.
86. Cocchiaro, J.L., et al., *Cytoplasmic lipid droplets are translocated into the lumen of the Chlamydia trachomatis parasitophorous vacuole*. *Proc Natl Acad Sci U S A*, 2008. **105**(27): p. 9379-84.
87. Dumoux, M., et al., *A Chlamydia effector recruits CEP170 to reprogram host microtubule organization*. *J Cell Sci*, 2015. **128**(18): p. 3420-34.
88. Hackstadt, T., et al., *The Chlamydia trachomatis IncA protein is required for homotypic vesicle fusion*. *Cellular microbiology*, 1999. **1**(2): p. 119-130.
89. Rzomp, K.A., A.R. Moorhead, and M.A. Scidmore, *The GTPase Rab4 interacts with Chlamydia trachomatis inclusion membrane protein CT229*. *Infect Immun*, 2006. **74**(9): p. 5362-73.
90. Rzomp, K.A., et al., *Rab GTPases are recruited to chlamydial inclusions in both a species-dependent and species-independent manner*. *Infect Immun*, 2003. **71**(10): p. 5855-70.
91. Rejman Lipinski, A., et al., *Rab6 and Rab11 regulate Chlamydia trachomatis development and golgin-84-dependent Golgi fragmentation*. *PLoS Pathog*, 2009. **5**(10): p. e1000615.
92. Derre, I., R. Swiss, and H. Agaisse, *The Lipid Transfer Protein CERT Interacts with the Chlamydia Inclusion Protein IncD and Participates to ER-Chlamydia Inclusion Membrane Contact Sites*. *PloS Pathogens*, 2011. **7**(6): p. e1002092.
93. Saka, H.A., et al., *Chlamydia trachomatis Infection Leads to Defined Alterations to the Lipid Droplet Proteome in Epithelial Cells*. *PLoS One*, 2015. **10**(4): p. e0124630.
94. Scidmore, M.A. and T. Hackstadt, *Mammalian 14-3-3beta associates with the Chlamydia trachomatis inclusion membrane via its interaction with IncG*. *Mol Microbiol*, 2001. **39**(6): p. 1638-50.
95. Verbeke, P., et al., *Recruitment of BAD by the Chlamydia trachomatis Vacuole Correlates with Host-Cell Survival*. *PLoS Pathogens*, 2006. **2**(5): p. e45.
96. Mirrashidi, K.M., et al., *Global Mapping of the Inc-Human Interactome Reveals that Retromer Restricts Chlamydia Infection*. *Cell Host Microbe*, 2015. **18**(1): p. 109-21.

97. Mital, J., et al., *Specific chlamydial inclusion membrane proteins associate with active Src family kinases in microdomains that interact with the host microtubule network*. Cellular Microbiology, 2010. **12**(9): p. 1235-1249.
98. Mital, J., et al., *Chlamydia trachomatis inclusion membrane protein CT850 interacts with the dynein light chain DYNLT1 (Tctex1)*. Biochem Biophys Res Commun, 2015. **462**(2): p. 165-70.
99. Lutter, E.I., et al., *Chlamydia trachomatis inclusion membrane protein CT228 recruits elements of the myosin phosphatase pathway to regulate release mechanisms*. Cell Reports, 2013. **3**(6): p. 1921-1931.
100. Mukhopadhyay, S., et al., *Identification of Chlamydia pneumoniae proteins in the transition from reticulate to elementary body formation*. Molecular & Cellular Proteomics, 2006. **5**(12): p. 2311-2318.
101. Zuck, M., et al., *Conservation of extrusion as an exit mechanism for Chlamydia*. Pathogens and Disease, 2016. **74**(7).
102. Wang, J., et al., *Altered protein secretion of Chlamydia trachomatis in persistently infected human endocervical epithelial cells*. Microbiology, 2011. **157**(Pt 10): p. 2759-2771.
103. Wyrick, P.B., *Chlamydia trachomatis persistence in vitro: an overview*. Journal of Infectious Diseases, 2010. **201** (Supplementary 2): p. S88-95.
104. Zhu, H., et al., *Persistent and acute chlamydial infections induce different structural changes in the Golgi apparatus*. International Journal of Medical Microbiology, 2014.
105. Belland, R.J., et al., *Transcriptome analysis of chlamydial growth during IFN-gamma-mediated persistence and reactivation*. Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(26): p. 15971-15976.
106. Ong, V.A., et al., *The protease inhibitor JO146 demonstrates a critical role for CtHtrA for Chlamydia trachomatis reversion from penicillin persistence*. Frontiers in Cellular and Infection Microbiology, 2013. **3**(100): p. 1-10.
107. Hayes, L.J., et al., *Extent and kinetics of genetic change in the omp1 gene of Chlamydia trachomatis in two villages with endemic trachoma*. J Infect Dis, 1995. **172**(1): p. 268-72.
108. Andreasen, A.A., et al., *Chlamydia trachomatis ompA variants in trachoma: what do they tell us?* PLoS Neglected Tropical Diseases, 2008. **2**(9): p. e306.
109. Hsieh, Y.H., et al., *Determinants of trachoma endemicity using Chlamydia trachomatis ompA DNA sequencing*. Microbes and Infection, 2001. **3**(6): p. 447-458.
110. Carlson, J.H., et al., *Polymorphisms in the Chlamydia trachomatis cytotoxin locus associated with ocular and genital isolates*. Infect Immun, 2004. **72**(12): p. 7063-7072.
111. Caldwell, H.D., et al., *Polymorphisms in Chlamydia trachomatis tryptophan synthase genes differentiate between genital and ocular isolates*. Journal of Clinical Investigation, 2003. **111**(11): p. 1757-1769.
112. Lutter, E.I., et al., *Phylogenetic analysis of Chlamydia trachomatis Tarp and correlation with clinical phenotype*. Infection and Immunity, 2010. **78**(9): p. 3678-388.
113. Gomes, J.P., et al., *Polymorphisms in the nine polymorphic membrane proteins of Chlamydia trachomatis across all serovars: evidence for serovar Da recombination and correlation with tissue tropism*. Journal of Bacteriology, 2006. **188**(1): p. 275-286.

114. Thomson, N.R., et al., *Chlamydia trachomatis*: genome sequence analysis of lymphogranuloma venereum isolates. *Genome Research*, 2008. **18**(1): p. 161-171.
115. Joseph, S.J., et al., *Interplay of recombination and selection in the genomes of Chlamydia trachomatis*. *Biology Direct*, 2011. **6**(28).
116. Kari, L., et al., *Pathogenic diversity among Chlamydia trachomatis ocular strains in nonhuman primates is affected by subtle genomic variations*. *Journal of Infectious Disease*, 2008. **197**(3): p. 449-456.
117. Rasmussen, S.J., et al., *Secretion of proinflammatory cytokines by epithelial cells in response to Chlamydia infection suggests a central role for epithelial cells in chlamydial pathogenesis*. *The Journal of Clinical Investigation*, 1997. **99**(1): p. 77-87.
118. Burton, M.J., et al., *Active trachoma is associated with increased conjunctival expression of IL17A and profibrotic cytokines*. *Infection and Immunity*, 2011. **79**(12): p. 4977-4983.
119. el-Asrar, A.M., et al., *Immunopathology of trachomatous conjunctivitis*. *British Journal of Ophthalmology*, 1989. **73**(4): p. 276-282.
120. Guzey, M., et al., *A survey of trachoma: the histopathology and the mechanism of progressive cicatrization of eyelid tissues*. *Ophthalmologica*, 2000. **214**(4): p. 277-284.
121. Natividad, A., et al., *Human conjunctival transcriptome analysis reveals the prominence of innate defense in Chlamydia trachomatis infection*. *Infection and Immunity*, 2010. **78**(11): p. 4895-4911.
122. Kunitomo, D. and R.C. Brunham, *Human Immune Response and Chlamydia trachomatis Infection*. *Review of Infectious Diseases*, 1985. **7**(5): p. 665-673.
123. Mabey, D.C.W., et al., *Expression of MHC class II antigens by conjunctival epithelial cells in trachoma: Implications concerning the pathogenesis of blinding disease*. *Journal of Clinical Pathology*, 1991. **44**(4): p. 285-289.
124. Mabey, D.C.W., et al., *Lymphocyte proliferative responses to chlamydial antigens in human chlamydial eye infections*. *Clinical and Experimental Immunology*, 1991. **86**(1): p. 37-42.
125. Bailey, R.L., et al., *Subjects Recovering from Human Ocular Chlamydial Infection Have Enhanced Lymphoproliferative Responses to Chlamydial Antigens Compared with Those of Persistently Diseased Controls*. *Infection and Immunity*, 1995. **63**(2): p. 389-392.
126. Holland, M.J., et al., *Synthetic peptides based on Chlamydia trachomatis antigens identify cytotoxic T lymphocyte responses in subjects from a trachoma-endemic population*. *Clinical and Experimental Immunology*, 1997. **107**(1): p. 44-49.
127. Bobo, L., et al., *Evidence for a Predominant Proinflammatory Conjunctival Cytokine Response in Individuals with Trachoma*. *Infection and Immunity*, 1996. **64**(8): p. 3273-3279.
128. Faal, N., *Conjunctival Immune Responses In Human Ocular Chlamydial Infections*, in *Medical Research Council, The Gambia*. 2011, Open University, UK.
129. Treharne, J.D., et al., *Antichlamydial antibody in tears and sera, and serotypes of Chlamydia trachomatis isolated from schoolchildren in Southern Tunisia*. *Br J Ophthalmol*, 1978. **62**(8): p. 509-15.
130. Collier, L.H., J. Sowa, and S. Sowa, *The serum and conjunctival antibody response to trachoma in Gambian children*. *Journal of Hygiene (Lond)*, 1972. **70**(4): p. 727-740.

131. Barenfanger, J. and A.B. MacDonald, *The role of immunoglobulin in the neutralization of trachoma infectivity*. Journal of Immunology, 1974. **113**(5): p. 1607-1617.
132. Caldwell, H.D. and L.J. Perry, *Neutralization of Chlamydia trachomatis Infectivity with Antibodies to the Major Outer Membrane Protein*. Infection and Immunity, 1982. **38**(2): p. 745-754.
133. Peterson, E.M., et al., *Effects of Antibody Isotype and Host Cell Type on In Vitro Neutralization of Chlamydia trachomatis*. Infection and Immunity, 1993. **61**(2): p. 498-503.
134. Crane, D.D., et al., *Chlamydia trachomatis polymorphic membrane protein D is a species-common pan-neutralizing antigen*. Proceedings of the National Academy of Sciences United States of America, 2006. **103**(6): p. 1894-1899.
135. Armitage, C.W., et al., *Divergent outcomes following transcytosis of IgG targeting intracellular and extracellular chlamydial antigens*. Immunology & Cell Biology, 2014.
136. Kari, L., et al., *Antibody signature of spontaneous clearance of Chlamydia trachomatis ocular infection and partial resistance against re-challenge in a nonhuman primate trachoma model*. PLoS Neglected Tropical Diseases, 2013. **7**(5): p. e2248.
137. Stephens, R.S., *The cellular paradigm of chlamydial pathogenesis*. Trends in Microbiology, 2003. **11**(1): p. 44-51.
138. Peeling, R.W., et al., *Antibody Response to the 60-kDa Chlamydial Heat-Shock Protein Is Associated with Scarring Trachoma*. Journal of Infectious Diseases, 1998. **177**(1): p. 256-259.
139. Hessel, T., et al., *Immune response to chlamydial 60-kilodalton heat shock protein in tears from Nepali trachoma patients*. Infection and Immunity, 2001. **69**(8): p. 4996-5000.
140. Skwor, T., et al., *Characterization of humoral immune responses to chlamydial HSP60, CPAF, and CT795 in inflammatory and severe trachoma*. Investigative Ophthalmology and Visual Science, 2010. **51**(10): p. 5128-5136.
141. Holland, M.J., et al., *Conjunctival Scarring in Trachoma Is Associated with Depressed Cell-Mediated Immune Responses to Chlamydial Antigens*. Journal of Infectious Diseases, 1993. **168**(6): p. 1528-1531.
142. Faal, N., et al., *Conjunctival FOXP3 expression in trachoma: do regulatory T cells have a role in human ocular Chlamydia trachomatis infection?* PLoS Medicine, 2006. **3**(8): p. e266.
143. Burton, M.J., et al., *Conjunctival transcriptome in scarring trachoma*. Infection and Immunity, 2011. **79**(1): p. 499-511.
144. Holland, M.J., et al., *Pathway-focused arrays reveal increased matrix metalloproteinase-7 (matrilysin) transcription in trachomatous trichiasis*. Investigative Ophthalmology and Visual Science, 2010. **51**(8): p. 3893-3902.
145. Derrick, T., et al., *Increased Epithelial Expression of CTGF and S100A7 with Elevated Subepithelial Expression of IL-1beta in Trachomatous Trichiasis*. PLoS Negl Trop Dis, 2016. **10**(6): p. e0004752.
146. Hu, V., et al., *Bacterial Infection in Scarring Trachoma*. Investigative Ophthalmology & Visual Science, 2011. **52**(5): p. 2181-2186.
147. Newhall, W.J., B. Batteiger, and R.B. Jones, *Analysis of the human serological response to proteins of Chlamydia trachomatis*. Infect Immun, 1982. **38**(3): p. 1181-9.
148. Ward, M.E., J.D. Treharne, and A. Murray, *Antigenic specificity of human antibody to chlamydia in trachoma and lymphogranuloma venereum*. J Gen Microbiol, 1986. **132**(6): p. 1599-610.

149. Caldwell, H.D., et al., *Tear and serum antibody response to Chlamydia trachomatis antigens during acute chlamydial conjunctivitis in monkeys as determined by immunoblotting*. Infect Immun, 1987. **55**(1): p. 93-8.
150. Davies, D.H., et al., *Profiling the humoral immune response to infection by using proteome microarrays: high-throughput vaccine and diagnostic antigen discovery*. Proceedings of the National Academy of Science of the United States of America, 2005. **102**(3): p. 547-552.
151. Doolan, D.L., et al., *Profiling humoral immune responses to P. falciparum infection with protein microarrays*. Proteomics, 2008. **8**(22): p. 4680-94.
152. Trieu, A., et al., *Sterile protective immunity to malaria is associated with a panel of novel P. falciparum antigens*. Mol Cell Proteomics, 2011. **10**(9): p. M111 007948.
153. Baum, E., et al., *Protein microarray analysis of antibody responses to Plasmodium falciparum in western Kenyan highland sites with differing transmission levels*. PLoS One, 2013. **8**(12): p. e82246.
154. Driguez, P., et al., *Schistosomiasis vaccine discovery using immunomics*. Parasit Vectors, 2010. **3**: p. 4.
155. Gaze, S., et al., *An immunomics approach to schistosome antigen discovery: antibody signatures of naturally resistant and chronically infected individuals from endemic areas*. PLoS Pathog, 2014. **10**(3): p. e1004033.
156. Kunnath-Velayudhan, S., et al., *Dynamic antibody responses to the Mycobacterium tuberculosis proteome*. Proc Natl Acad Sci U S A, 2010. **107**(33): p. 14703-8.
157. Nandakumar, S., et al., *O-mannosylation of the Mycobacterium tuberculosis adhesin Apa is crucial for T cell antigenicity during infection but is expendable for protection*. PLoS Pathog, 2013. **9**(10): p. e1003705.
158. Khan, I.H., et al., *Plasma antibody profiles as diagnostic biomarkers for tuberculosis*. Clin Vaccine Immunol, 2011. **18**(12): p. 2148-53.
159. Yang, H., et al., *A novel B-cell epitope identified within Mycobacterium tuberculosis CFP10/ESAT-6 protein*. PLoS One, 2013. **8**(1): p. e52848.
160. Kunnath-Velayudhan, S., et al., *Proteome-scale antibody responses and outcome of Mycobacterium tuberculosis infection in nonhuman primates and in tuberculosis patients*. J Infect Dis, 2012. **206**(5): p. 697-705.
161. Sharma, J., et al., *Profiling of human antibody responses to Chlamydia trachomatis urogenital tract infection using microplates arrayed with 156 chlamydial fusion proteins*. Infection and Immunity, 2006. **74**(3): p. 1490-1499.
162. Lu, C., et al., *Genome-wide identification of Chlamydia trachomatis antigens associated with trachomatous trichiasis*. Investigative Ophthalmology and Visual Science, 2012. **53**(6): p. 2551-2559.
163. Mital, J., et al., *Role for Chlamydial Inclusion Membrane Proteins in Inclusion Membrane Structure and Biogenesis*. PloS One, 2013. **8**(5): p. e63426.
164. Raran-Kurussi, S. and D.S. Waugh, *The ability to enhance the solubility of its fusion partners is an intrinsic property of maltose-binding protein but their folding is either spontaneous or chaperone-mediated*. PLoS One, 2012. **7**(11): p. e49589.
165. Terpe, K., *Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems*. Applied Microbiology and Biotechnology, 2003. **60**(5): p. 523-533.
166. Stekel, D., *Microarray Bioinformatics*. 2003, Cambridge, UK: Cambridge University Press. 280.

167. Budrys, N.M., et al., *Chlamydia trachomatis* antigens recognized in women with tubal factor infertility, normal fertility, and acute infection. *Obstetrics & Gynecology*, 2012. **119**(5): p. 1009-1016.
168. Rodgers, A.K., et al., *Genome-wide identification of Chlamydia trachomatis* antigens associated with tubal factor infertility. *Fertility and Sterility*, 2011. **96**(3): p. 715-721.
169. Wang, J., et al., *A genome-wide profiling of the humoral immune response to Chlamydia trachomatis* infection reveals vaccine candidate antigens expressed in humans. *Journal of Immunology*, 2010. **185**(3): p. 1670-1680.
170. Follmann, F., et al., *Antigenic profiling of a Chlamydia trachomatis gene-expression library*. *Journal of Infectious Diseases*, 2008. **197**(6): p. 897-905.
171. Finco, O., et al., *Approach to discover T- and B-cell antigens of intracellular pathogens applied to the design of Chlamydia trachomatis vaccines*. *Proceedings of the National Academy of Science of the United States of America*, 2011. **108**(24): p. 9969-9974.
172. Coler, R.N., et al., *Identification and characterization of novel recombinant vaccine antigens for immunization against genital Chlamydia trachomatis*. *FEMS Immunology and Medical Microbiology*, 2009. **55**(2): p. 258-270.
173. Nichols, R.L., et al., *Studies on Trachoma*. *American Journal of Tropical Medicine and Hygiene*, 1968. **15**(4): p. 639-647.
174. Collier, L.H., *Trachoma and allied infections*. *Transactions of the Ophthalmological Societies of the United Kingdom*, 1961. **81**: p. 351-365.
175. Grayston, J.T., et al., *Field studies of protection from infection by experimental trachoma virus vaccine in preschool-aged children on Taiwan*. *Proceedings of the Society for Experimental Biology and Medicine*, 1963. **112**: p. 589-95.
176. Clements, C., et al., *Long term follow-up study of a trachoma vaccine trial in villages of Northern India*. *Am J Ophthalmol*, 1979. **87**(3): p. 350-3.
177. Batteiger, B.E., *The Major Outer Membrane Protein of a Single Chlamydia trachomatis Serovar Can Possess More than One Serovar-Specific Epitope*. *Infection and Immunity*, 1996. **64**(2): p. 542-547.
178. Cambridge, C.D., et al., *Formulation, characterization, and expression of a recombinant MOMP Chlamydia trachomatis DNA vaccine encapsulated in chitosan nanoparticles*. *International Journal of Nanomedicine*, 2013. **8**: p. 1759-1771.
179. Cheng, C., et al., *A TLR2 agonist is a more effective adjuvant for a Chlamydia major outer membrane protein vaccine than ligands to other TLR and NOD receptors*. *Vaccine*, 2011. **29**(38): p. 6641-6649.
180. Cheng, C., et al., *A vaccine formulated with a combination of TLR-2 and TLR-9 adjuvants and the recombinant major outer membrane protein elicits a robust immune response and significant protection against a Chlamydia muridarum challenge*. *Microbes and Infection*, 2013.
181. Kari, L., et al., *Chlamydia trachomatis native major outer membrane protein induces partial protection in nonhuman primates: implication for a trachoma transmission-blocking vaccine*. *Journal of Immunology*, 2009. **182**(12): p. 8063-8070.
182. Ou, C., et al., *Evaluation of an ompA-based phage-mediated DNA vaccine against Chlamydia abortus in piglets*. *International Immunopharmacology*, 2013. **16**(4): p. 505-510.
183. Dixit, S., et al., *Poly (lactic acid)-poly (ethylene glycol) nanoparticles provide sustained delivery of a Chlamydia trachomatis recombinant MOMP peptide and potentiate systemic adaptive immune responses in mice*. *Nanomedicine*, 2014.

184. Fairley, S.J., et al., *Chlamydia trachomatis recombinant MOMP encapsulated in PLGA nanoparticles triggers primarily T helper 1 cellular and antibody immune responses in mice: a desirable candidate nanovaccine*. International Journal of Nanomedicine, 2013. **8**: p. 2085-2099.
185. Su, H., M. Parnell, and H.D. Caldwell, *Protective efficacy of a parenterally administered MOMP-derived synthetic oligopeptide vaccine in a murine model of Chlamydia trachomatis genital tract infection: serum neutralizing IgG antibodies do not protect against chlamydial genital tract infection*. Vaccine, 1995. **13**(11): p. 1023-1032.
186. Brunham, R.C. and J. Rey-Ladino, *Immunology of Chlamydia infection: implications for a Chlamydia trachomatis vaccine*. Nature Reviews Immunology, 2005. **5**(2): p. 149-161.
187. Inic-Kanada, A., et al., *Delivery of a Chlamydial Adhesin N-PmpC Subunit Vaccine to the Ocular Mucosa Using Particulate Carriers*. PLoS One, 2015. **10**(12): p. e0144380.
188. Olsen, A.W., et al., *Protection Against Chlamydia trachomatis Infection and Upper Genital Tract Pathological Changes by Vaccine-Promoted Neutralizing Antibodies Directed to the VD4 of the Major Outer Membrane Protein*. Journal of Infectious Diseases, 2015. **4**(13): p. 978-989.
189. Stary, G., et al., *A mucosal vaccine against Chlamydia trachomatis generates two waves of protective memory T cells*. Science, 2015. **348**(6241): p. aaa8205.
190. Pal, S., O.V. Tatarenkova, and L.M. de la Maza, *A vaccine formulated with the major outer membrane protein can protect C3H/HeN, a highly susceptible strain of mice, from a Chlamydia muridarum genital challenge*. Immunology, 2015.
191. Boje, S., et al., *A multi-subunit Chlamydia vaccine inducing neutralizing antibodies and strong IFN-gamma(+) CMI responses protects against a genital infection in minipigs*. Immunology and Cell Biology, 2016. **94**(2): p. 185-195.
192. Paes, W., et al., *Recombinant polymorphic membrane protein D in combination with a novel, second-generation lipid adjuvant protects against intra-vaginal Chlamydia trachomatis infection in mice*. Vaccine, 2016. **34**(35): p. 4123-31.
193. Yu, H., et al., *Chlamydia muridarum T cell antigens and adjuvants that induce protective immunity in mice*. Infection and Immunity, 2012. **80**(4): p. 1510-1518.
194. O'Meara, C.P., D.W. Andrew, and K.W. Beagley, *The Mouse Model of Chlamydia Genital Tract Infection: A Review of Infection, Disease, Immunity and Vaccine Development*. Current Molecular Medicine, 2014. **14**(3): p. 396-421.
195. Starnbach, M.N., et al., *An Inclusion Membrane Protein from Chlamydia trachomatis Enters the MHC Class I Pathway and Stimulates a CD8⁺ T Cell Response*. The Journal of Immunology, 2003. **171**(9): p. 4742-4749.
196. Feinstein, H.E., et al., *Long-term stability of a vaccine formulated with the amphipol-trapped major outer membrane protein from Chlamydia trachomatis*. Journal of Membrane Biology, 2014. **247**(9-10): p. 1053-1065.
197. Griffiths, K.L. and S.A. Khader, *Novel vaccine approaches for protection against intracellular pathogens*. Current Opinions in Immunology, 2014. **28**: p. 58-63.
198. Olsen, A.W., P. Andersen, and F. Follmann, *Characterization of protective immune responses promoted by human antigen targets in a urogenital Chlamydia trachomatis mouse model*. Vaccine, 2014. **32**(6): p. 685-692.
199. Nogueira, C.V., et al., *Protective Immunity against Chlamydia trachomatis Can Engage Both CD4⁺ and CD8⁺ T Cells and Bridge the Respiratory and Genital Mucosae*. Journal of Immunology, 2015.

200. Olivares-Zavaleta, N., et al., *CD8+ T Cells Define an Unexpected Role in Live-Attenuated Vaccine Protective Immunity against Chlamydia trachomatis Infection in Macaques*. The Journal of Immunology, 2014. **192**(10): p. 4648-4654.
201. Kari, L., et al., *A live-attenuated chlamydial vaccine protects against trachoma in nonhuman primates*. Journal of Experimental Medicine, 2011. **208**(11): p. 2217-2223.
202. Nishihara, K., et al., *Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in Escherichia coli*. Appl Environ Microbiol, 1998. **64**(5): p. 1694-9.
203. Nishihara, K., et al., *Overexpression of trigger factor prevents aggregation of recombinant proteins in Escherichia coli*. Appl Environ Microbiol, 2000. **66**(3): p. 884-9.
204. Rahman, K.S., et al., *Defining species-specific immunodominant B cell epitopes for molecular serology of Chlamydia species*. Clinical and Vaccine Immunology, 2015. **22**(5): p. 539-552.
205. Patton, D.L., et al., *Whole genome identification of C. trachomatis immunodominant antigens after genital tract infections and effect of antibiotic treatment of pigtailed macaques*. Journal of Proteomics, 2014. **108**: p. 99-109.
206. Kroll, T.C. and S. Wolfl, *Ranking: a closer look on globalisation methods for normalisation of gene expression arrays*. Nucleic Acids Research, 2002. **30**(11): p. e50.
207. Burton, M.J., et al., *Cytokine and fibrogenic gene expression in the conjunctivas of subjects from a Gambian community where trachoma is endemic*. Infection and Immunity, 2004. **72**(12): p. 7352-7356.
208. *WHO simplified trachoma grading system*. Community Eye Health, 2004. **17**(52): p. 68-68.
209. Magurran, A.E., *Assumptions of biodiversity measurement*, in *Measuring Biological Diversity*. 2004, Blackwell Science: MA, USA. p. 256.
210. Gotelli, N.J. and A. Chao, *Measuring and Estimating Species Richness, Species Diversity, and Biotic Similarity from Sampling Data*, in *Encyclopedia of Biodiversity*, S. Levin, Editor. 2013, Academic Press: Waltham, MA. p. 195-211.
211. Morris, E.K., et al., *Choosing and using diversity indices: insights for ecological applications from the German Biodiversity Exploratories*. Ecology and Evolution, 2014. **4**(18): p. 3514-3524.
212. Belland, R.J., et al., *Genomic transcriptional profiling of the developmental cycle of Chlamydia trachomatis*. Proceedings of the National Academy of Science of the United States of America, 2003. **100**(14): p. 8478-8483.
213. Nicholson, T.L., et al., *Global Stage-Specific Gene Regulation during the Developmental Cycle of Chlamydia trachomatis*. Journal of Bacteriology, 2003. **185**(10): p. 3179-3189.
214. Yu, C.S., C.J. Lin, and J.K. Hwang, *Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions*. Protein Science, 2004. **13**(5): p. 1402-1406.
215. Yu, N.Y., et al., *PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes*. Bioinformatics, 2010. **26**(13): p. 1608-1615.
216. Goldberg, T., et al., *LocTree3 prediction of localization*. Nucleic Acids Research, 2014. **42**(Web Server issue): p. W350-5.

217. Shen, H. and J.J. Chou, *MemBrain: improving the accuracy of predicting transmembrane helices*. PLoS One, 2008. **3**(6): p. e2399.
218. Sonnhammer, E.L., G. von Heijne, and A. Krogh, *A hidden Markov model for predicting transmembrane helices in protein sequences*. Proc Int Conf Intell Syst Mol Biol, 1998. **6**: p. 175-82.
219. Zaki, N., S. Bouktif, and S. Lazarova-Molnar, *A combination of compositional index and genetic algorithm for predicting transmembrane helical segments*. PLoS One, 2011. **6**(7): p. e21821.
220. Andreatta, M. and M. Nielsen, *Gapped sequence alignment using artificial neural networks: application to the MHC class I system*. Bioinformatics, 2016. **32**(4): p. 511-7.
221. Moutaftsi, M., et al., *A consensus epitope prediction approach identifies the breadth of murine T(CD8+)-cell responses to vaccinia virus*. Nat Biotechnol, 2006. **24**(7): p. 817-9.
222. Nielsen, M. and O. Lund, *NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction*. BMC Bioinformatics, 2009. **10**: p. 296.
223. Larsen, J.E., O. Lund, and M. Nielsen, *Improved method for predicting linear B-cell epitopes*. Immunome Res, 2006. **2**: p. 2.
224. Saha, S. and G.P. Raghava, *Prediction of continuous B-cell epitopes in an antigen using recurrent neural network*. Proteins, 2006. **65**(1): p. 40-8.
225. Chou, P.Y. and G.D. Fasman, *Prediction of the secondary structure of proteins from their amino acid sequence*. Adv Enzymol Relat Areas Mol Biol, 1978. **47**: p. 45-148.
226. Emini, E.A., et al., *Induction of hepatitis A virus-neutralizing antibody by a virus-specific synthetic peptide*. J Virol, 1985. **55**(3): p. 836-9.
227. Kolaskar, A.S. and P.C. Tongaonkar, *A semi-empirical method for prediction of antigenic determinants on protein antigens*. FEBS Lett, 1990. **276**(1-2): p. 172-4.
228. Parker, J.M., D. Guo, and R.S. Hodges, *New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites*. Biochemistry, 1986. **25**(19): p. 5425-32.
229. Greenbaum, J.A., et al., *Towards a consensus on datasets and evaluation metrics for developing B-cell epitope prediction tools*. J Mol Recognit, 2007. **20**(2): p. 75-82.
230. Ponomarenko, J.V. and M.H.V. van Regenmortel, *B-Cell Epitope Prediction*, in *Structural Bioinformatics*, J. Gu and P.E. Bourne, Editors. 2009, John Wiley & Sons: Hoboken, NJ. p. 849-880.
231. Conway, D.J., et al., *HLA Class I and II Polymorphisms and Trachomatous Scarring in a Chlamydia trachomatis-Endemic Population*. The Journal of Infectious Diseases, 1996. **174**(3): p. 643-646.
232. Conway, D.J., et al., *Scarring Trachoma Is Associated with Polymorphism in the Tumor Necrosis Factor Alpha (TNF-alpha) Gene Promoter and with Elevated TNF-alpha Levels in Tear Fluid*. Infection and Immunity, 1997. **65**(3): p. 1003-1006.
233. Gall, A., et al., *Systemic effector and regulatory immune responses to chlamydial antigens in trachomatous trichiasis*. Frontiers in Microbiology, 2011. **2**(10).
234. Wills, G.S., et al., *Pgp3 antibody enzyme-linked immunosorbent assay, a sensitive and specific assay for seroepidemiological analysis of Chlamydia*

- trachomatis* infection. Clinical and Vaccine Immunology, 2009. **16**(6): p. 835-843.
235. Faal, N., et al., *Temporal cytokine gene expression patterns in subjects with trachoma identify distinct conjunctival responses associated with infection*. Clinical and Experimental Immunology, 2005. **142**(2): p. 347-353.
 236. Roberts, C.H., et al., *Development and Evaluation of a Next-Generation Digital PCR Diagnostic Assay for Ocular Chlamydia trachomatis Infections*. Journal of Clinical Microbiology, 2013. **51**(7): p. 2195-2203.
 237. Goodier, M.R., et al., *Rapid NK cell differentiation in a population with near-universal human cytomegalovirus infection is attenuated by NKG2C deletions*. Blood, 2014. **124**(14): p. 2213-22.
 238. Butcher, R.M., et al., *Low Prevalence of Conjunctival Infection with Chlamydia trachomatis in a Treatment-Naive Trachoma-Endemic Region of the Solomon Islands*. PLoS Negl Trop Dis, 2016. **10**(9): p. e0004863.
 239. Martin, D.L., et al., *Serology for Post-Elimination Surveillance in Trachoma Programs Following Cessation of Antibiotic Mass Drug Administration*. 2014.
 240. Last, A.R., et al., *Plasmid Copy Number and Disease Severity in Naturally Occurring Ocular Chlamydia trachomatis Infection*. Journal of Clinical Microbiology, 2014. **52**(1): p. 324-327.
 241. Skipp, P., et al., *Shotgun proteomic analysis of Chlamydia trachomatis*. Proteomics, 2005. **5**(6): p. 1558-73.
 242. Seth-Smith, H.M., et al., *Co-evolution of genomes and plasmids within Chlamydia trachomatis and the emergence in Sweden of a new variant strain*. BMC Genomics, 2009. **10**: p. 239.
 243. Andrews, S. (2010) *FastQC: a quality control tool for high throughput sequence data*. **Volume**,
 244. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform*. Bioinformatics, 2009. **25**(14): p. 1754-60.
 245. Li, H., et al., *The Sequence Alignment/Map format and SAMtools*. Bioinformatics, 2009. **25**(16): p. 2078-9.
 246. Danecek, P., et al., *The variant call format and VCFtools*. Bioinformatics, 2011. **27**(15): p. 2156-8.
 247. Camacho, C., et al., *BLAST+: architecture and applications*. BMC Bioinformatics, 2009. **10**: p. 421.
 248. Edgar, R.C., *MUSCLE: a multiple sequence alignment method with reduced time and space complexity*. BMC Bioinformatics, 2004. **5**: p. 113.
 249. Gouy, M., S. Guindon, and O. Gascuel, *SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building*. Mol Biol Evol, 2010. **27**(2): p. 221-4.
 250. Kearse, M., et al., *Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data*. Bioinformatics, 2012. **28**(12): p. 1647-9.
 251. Vilella, A.J., et al., *VariScan: Analysis of evolutionary patterns from large-scale DNA sequence polymorphism data*. Bioinformatics, 2005. **21**(11): p. 2791-3.
 252. Holsinger, K.E., *Tajima's D, Fu's F_s, Fay and Wu's H, and Zeng et al.'s E*. 2012, Stanford: California, USA. p. 1-7.
 253. Tajima, F., *Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism*. Genetics, 1989. **123**(3): p. 585-595.
 254. Fu, Y.-X., *Statistical Tests of Neutrality of Mutations Against Population Growth, Hitchhiking and Background Selection*. Genetics, 1997. **147**(2): p. 915-925.

255. Fay, J.C. and C.-I. Wu, *Hitchhiking Under Positive Darwinian Selection*. Genetics, 2000. **155**(3): p. 1405-1413.
256. Voight, B.F., et al., *A map of recent positive selection in the human genome*. PLoS Biol, 2006. **4**(3): p. e72.
257. Szpiech, Z.A. and R.D. Hernandez, *selscan: an efficient multithreaded program to perform EHH-based scans for positive selection*. Mol Biol Evol, 2014. **31**(10): p. 2824-7.
258. Aeberhard, L., et al., *The Proteome of the Isolated Chlamydia trachomatis Containing Vacuole Reveals a Complex Trafficking Platform Enriched for Retromer Components*. PLoS Pathogens, 2015. **11**(6): p. e1004883.
259. Huang, D.W., B.T. Sherman, and R.A. Lempicki, *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources*. Nat. Protocols, 2008. **4**(1): p. 44-57.
260. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium*. Nat Genet, 2000. **25**(1): p. 25-9.
261. Finn, R.D., et al., *The Pfam protein families database: towards a more sustainable future*. Nucleic Acids Res, 2016. **44**(D1): p. D279-85.
262. Hunter, S., et al., *InterPro: the integrative protein signature database*. Nucleic Acids Res, 2009. **37**(Database issue): p. D211-5.
263. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes*. Nucleic Acids Res, 2000. **28**(1): p. 27-30.
264. Croft, D., et al., *The Reactome pathway knowledgebase*. Nucleic Acids Res, 2014. **42**(Database issue): p. D472-7.
265. Szklarczyk, D., et al., *STRING v10: protein-protein interaction networks, integrated over the tree of life*. Nucleic Acids Res, 2015. **43**(Database issue): p. D447-52.
266. Ward, M., et al., *Persisting inapparent chlamydial infection in a trachoma endemic community in The Gambia*. Scandinavian Journal of Infectious Diseases. Supplementum, 1990. **69**: p. 137-148.
267. Bailey, R.L., et al., *Polymerase chain reaction for the detection of ocular chlamydial infection in trachoma-endemic communities*. Journal of Infectious Diseases, 1994. **170**(3): p. 709-712.
268. Bobo, L.D., et al., *Severe disease in children with trachoma is associated with persistent Chlamydia trachomatis infection*. J Infect Dis, 1997. **176**(6): p. 1524-30.
269. Treharne, J.D., *The microbial epidemiology of trachoma*. Int Ophthalmol, 1988. **12**(1): p. 25-9.
270. Treharne, J.D., *The community epidemiology of trachoma*. Reviews in Infectious Diseases, 1985. **7**(6): p. 760-764.
271. Taylor, H.R., et al., *The natural history of endemic trachoma: a longitudinal study*. The American Journal of Tropical Medicine and Hygiene, 1992. **46**(5): p. 552-559.
272. West, S.K., et al., *Risk factors for constant, severe trachoma among preschool children in Kongwa, Tanzania*. Am J Epidemiol, 1996. **143**(1): p. 73-8.
273. Smith, A., et al., *OmpA genotypic evidence for persistent ocular Chlamydia trachomatis infection in Tanzanian village women*. Ophthalmic Epidemiology, 2001. **8**(2-3): p. 127-135.
274. Beatty, W.L., G.I. Byrne, and R.P. Morrison, *Repeated and persistent infection with Chlamydia and the development of chronic inflammation and disease*. Trends Microbiol, 1994. **2**(3): p. 94-8.
275. Tseng, C.T. and R.G. Rank, *Role of NK cells in early host response to chlamydial genital infection*. Infect Immun, 1998. **66**(12): p. 5867-75.

276. Stagg, A.J., et al., *Acquisition of chlamydial antigen by dendritic cells and monocytes*. Advances in Experimental Medicine and Biology, 1993. **329**: p. 581-586.
277. Yang, X. and R. Brunham, *Gene Knockout B Cell-Deficient Mice Demonstrate That B Cells Play an Important Role in the Initiation of T Cell Responses to Chlamydia trachomatis (Mouse Pneumonitis) Lung Infection*. The Journal of Immunology, 1998. **161**(3): p. 1439-1446.
278. Moore, T., et al., *Fc Receptor Regulation of Protective Immunity Against Chlamydia trachomatis*. Immunology, 2002. **105**(2): p. 213-221.
279. Mabey, D.C., et al., *A longitudinal study of trachoma in a Gambian village: implications concerning the pathogenesis of chlamydial infection*. Epidemiology and Infection, 1992. **108**(2): p. 343-351.
280. Bailey, R.L., et al., *The influence of local antichlamydial antibody on the acquisition and persistence of human ocular chlamydial infection: IgG antibodies are not protective*. Epidemiology and Infection, 1993. **111**(2): p. 315-324.
281. Davies, D.H., et al., *Profiling the humoral immune response to infection by using proteome microarrays: high-throughput vaccine and diagnostic antigen discovery*. Proceedings of the National Academy of Sciences United States of America, 2005. **102**(3): p. 547-552.
282. Liang, L. and P.L. Felgner, *A systems biology approach for diagnostic and vaccine antigen discovery in tropical infectious diseases*. Current Opinion in Infectious Diseases, 2015. **28**(5): p. 438-445.
283. Helb, D.A., et al., *Novel serologic biomarkers provide accurate estimates of recent Plasmodium falciparum exposure for individuals and communities*. Proceedings of the National Academy of Sciences 2015.
284. Bonilla-Rosso, G., et al., *Understanding microbial community diversity metrics derived from metagenomes: performance evaluation using simulated data sets*. FEMS Microbiol Ecol, 2012. **82**(1): p. 37-49.
285. Comanducci, M., et al., *Humoral immune response to plasmid protein pgp3 in patients with Chlamydia trachomatis infection*. Infect Immun, 1994. **62**(12): p. 5491-7.
286. Goodhew, E.B., et al., *CT694 and pgp3 as serological tools for monitoring trachoma programs*. PLoS Neglected Tropical Diseases, 2012. **6**(11): p. e1873.
287. Bard, J. and D. Levitt, *Chlamydia trachomatis stimulates human peripheral blood B lymphocytes to proliferate and secrete polyclonal immunoglobulins in vitro*. Infect Immun, 1984. **43**(1): p. 84-92.
288. Bernkopf, H., J. Orfila, and B. Maythar, *Fluorescent antibodies in the fluid of the conjunctival sac of trachoma patients*. Nature, 1966. **209**(5024): p. 725-6.
289. Levitt, D. and R. Corlett, *Patterns of immunoenhancement and suppression induced by Chlamydia trachomatis in vivo and in vitro*. J Immunol, 1988. **140**(1): p. 273-6.
290. Nichols, R.L., et al., *Immunity to chlamydial infections of the eye. VI. Homologous neutralization of trachoma infectivity for the owl monkey conjunctivae by eye secretions from humans with trachoma*. J Infect Dis, 1973. **127**(4): p. 429-32.
291. Peeling, R., I.W. Maclean, and R.C. Brunham, *In vitro neutralization of Chlamydia trachomatis with monoclonal antibody to an epitope on the major outer membrane protein*. Infect Immun, 1984. **46**(2): p. 484-8.
292. Manz, R.A., et al., *Maintenance of serum antibody levels*. Annu Rev Immunol, 2005. **23**: p. 367-86.

293. Sather, D.N., et al., *Factors associated with the development of cross-reactive neutralizing antibodies during human immunodeficiency virus type 1 infection*. J Virol, 2009. **83**(2): p. 757-69.
294. Doria-Rose, N.A., et al., *Breadth of human immunodeficiency virus-specific neutralizing activity in sera: clustering analysis and association with clinical variables*. J Virol, 2010. **84**(3): p. 1631-6.
295. Kwong, P.D., J.R. Mascola, and G.J. Nabel, *Broadly neutralizing antibodies and the search for an HIV-1 vaccine: the end of the beginning*. Nat Rev Immunol, 2013. **13**(9): p. 693-701.
296. Bromuro, C., et al., *Interplay between protective and inhibitory antibodies dictates the outcome of experimentally disseminated Candidiasis in recipients of a Candida albicans vaccine*. Infection and Immunity, 2002. **70**(10): p. 5462-5470.
297. Lundemose, A.G., et al., *Chlamydia trachomatis Mip-like protein*. Mol Microbiol, 1992. **6**(17): p. 2539-48.
298. Goodall, J.C., et al., *Identification of Chlamydia trachomatis antigens recognized by human CD4+ T lymphocytes by screening an expression library*. European Journal of Immunology, 2001. **31**(5): p. 1513-1522.
299. Stallmann, S. and J.H. Hegemann, *The Chlamydia trachomatis Ctad1 invasin exploits the human integrin beta1 receptor for host cell entry*. Cell Microbiol, 2016. **18**(5): p. 761-75.
300. Abdelrahman, Y.M. and R.J. Belland, *The chlamydial developmental cycle*. FEMS Microbiology Reviews, 2005. **29**(5): p. 949-959.
301. Dehoux, P., et al., *Multi-genome identification and characterization of chlamydiae-specific type III secretion substrates: the Inc proteins*. BMC Genomics, 2011. **12**(109).
302. Giles, D.K., et al., *Ultrastructural analysis of chlamydial antigen-containing vesicles everting from the Chlamydia trachomatis inclusion*. Microbes Infect, 2006. **8**(6): p. 1579-91.
303. Frohlich, K., et al., *Isolation of Chlamydia trachomatis and membrane vesicles derived from host and bacteria*. J Microbiol Methods, 2012. **91**(2): p. 222-30.
304. Frohlich, K.M., et al., *Membrane vesicle production by Chlamydia trachomatis as an adaptive response*. Front Cell Infect Microbiol, 2014. **4**: p. 73.
305. Doolan, D.L., C. Dobano, and J.K. Baird, *Acquired immunity to malaria*. Clin Microbiol Rev, 2009. **22**(1): p. 13-36, Table of Contents.
306. Swanson, K.A., et al., *Chlamydia trachomatis polymorphic membrane protein D is an oligomeric autotransporter with a higher-order structure*. Infection and Immunity, 2009. **77**(1): p. 508-516.
307. Feher, V.A., et al., *A 3-dimensional trimeric beta-barrel model for Chlamydia MOMP contains conserved and novel elements of Gram-negative bacterial porins*. PLoS One, 2013. **8**(7): p. e68934.
308. West, S.K., et al., *Progression of active trachoma to scarring in a cohort of Tanzanian children*. Ophthalmic Epidemiology, 2001. **8**(2-3): p. 137-144.
309. Burton, M.J., et al., *Pathogenesis of progressive scarring trachoma in Ethiopia and Tanzania and its implications for disease control: two cohort studies*. PLoS Neglected Tropical Diseases, 2015. **9**(5): p. e0003763.
310. Ramadhani, A.M., et al., *Blinding Trachoma: Systematic Review of Rates and Risk Factors for Progressive Disease*. PLoS Neglected Tropical Diseases, 2016. **10**(8): p. e0004859.
311. Burton, M.J., et al., *The long-term natural history of trachomatous trichiasis in the Gambia*. Investigative Ophthalmology and Visual Science, 2006. **47**(3): p. 847-852.

312. Burton, M.J., et al., *Conjunctival Chlamydial 16S Ribosomal RNA Expression in Trachoma: Is Chlamydial Metabolic Activity Required for Disease to Develop?* Clinical Infectious Diseases, 2006. **42**(4): p. 463-370.
313. Blodi, B.A., K.A. Byrne, and K.F. Tabbara, *Goblet cell population among patients with inactive trachoma*. Int Ophthalmol, 1988. **12**(1): p. 41-5.
314. Guzey, M., et al., *The treatment of severe trichomatous dry eye with canalicular silicone plugs*. Eye (London, England), 2001. **15**(Pt 3): p. 297-303.
315. Lucena, A., et al., *Upper eyelid entropion and dry eye in cicatricial trachoma without trichiasis*. Arquivos Brasileiros de Oftalmologia, 2012. **75**(6): p. 420-422.
316. Skwor, T.A., et al., *Role of secreted conjunctival mucosal cytokine and chemokine proteins in different stages of trichomatous disease*. PLoS Neglected Tropical Diseases, 2008. **2**(7): p. e264.
317. Hu, V.H., et al., *In vivo confocal microscopy in scarring trachoma*. Ophthalmology, 2011. **118**(11): p. 2138-46.
318. Holland, M.J., *A Study of Cell Mediate Immune (CMI) Responses to Chlamydia trachomatis in Humans*, in *Department of Clinical Sciences*. 1993, London School of Hygiene and Tropical Medicine: London.
319. Holland, M.J., et al., *T helper type-1 (Th1)/Th2 profiles of peripheral blood mononuclear cells (PBMC); responses to antigens of Chlamydia trachomatis in subjects with severe trichomatous scarring*. Clinical and Experimental Immunology, 1996. **105**(3): p. 429-435.
320. Rodgers, A.K., et al., *Association of tubal factor infertility with elevated antibodies to Chlamydia trachomatis caseinolytic protease P*. American Journal of Obstetrics and Gynecology, 2010. **203**(5): p. 494-494.
321. Surana, A., V. Rastogi, and P.S. Nirwan, *Association of the Serum Anti-chlamydial Antibodies with Tubal Infertility*. Journal of Clinical and Diagnostic Research, 2012. **6**(10): p. 1692-1694.
322. Scidmore-Carlson, M.A., et al., *Identification and characterization of a Chlamydia trachomatis early operon encoding four novel inclusion membrane proteins*. Mol Microbiol, 1999. **33**(4): p. 753-65.
323. Fadel, S. and A. Eley, *Chlamydia trachomatis OmcB protein is a surface-exposed glycosaminoglycan-dependent adhesin*. Journal of Medical Microbiology, 2007. **56**(Pt 1): p. 15-22.
324. Sun, G., et al., *Structural and functional analyses of the major outer membrane protein of Chlamydia trachomatis*. Journal of Bacteriology, 2007. **189**(17): p. 6222-6235.
325. Raulston, J.E., et al., *Localization of Chlamydia trachomatis heat shock proteins 60 and 70 during infection of a human endometrial epithelial cell line in vitro*. Infection and Immunity, 1998. **66**(5): p. 2323-2329.
326. Qi, M., et al., *A Chlamydia trachomatis OmcB C-terminal fragment is released into the host cell cytoplasm and is immunogenic in humans*. Infection and Immunity, 2011. **79**(6): p. 2193-2203.
327. Betts, H.J., et al., *Bioinformatic and biochemical evidence for the identification of the type III secretion system needle protein of Chlamydia trachomatis*. Journal of Bacteriology, 2008. **190**(5): p. 1680-1690.
328. Spedding, L., *Novel Effector Protein Secretion and Transcriptional Regulation of the Type Three Secretion System in Chlamydia trachomatis*, in *Molecular Biosciences*. 2009, University of Kansas. p. 116.
329. Liu, X., et al., *Identification of Chlamydia trachomatis outer membrane complex proteins by differential proteomics*. Journal of Bacteriology, 2010. **192**(11): p. 2852-2860.

330. Sharma, J., et al., *Human antibody responses to a Chlamydia-secreted protease factor*. Infection and Immunity, 2004. **72**(12): p. 7164-7171.
331. Bartolini, E., et al., *Recombinant outer membrane vesicles carrying Chlamydia muridarum HtrA induce antibodies that neutralize chlamydial infection in vitro*. Journal of Extracellular Vesicles, 2013. **2**.
332. Almeida, F., et al., *Polymorphisms in inc proteins and differential expression of inc genes among Chlamydia trachomatis strains correlate with invasiveness and tropism of lymphogranuloma venereum isolates*. Journal of Bacteriology, 2012. **194**(23): p. 6574-6585.
333. Tang, L., et al., *Chlamydia-secreted protease CPAF degrades host antimicrobial peptides*. Microbes and Infection, 2015. **17**(6): p. 402-408.
334. Yang, Z., et al., *Neutralizing antichlamydial activity of complement by chlamydia-secreted protease CPAF*. Microbes and Infection, 2016. **18**(11): p. 669-674.
335. Porankiewicz, J., J. Wang, and A.K. Clarke, *New insights into the ATP-dependent Clp protease: Escherichia coli and beyond*. Molecular Microbiology, 1999. **32**(3): p. 449-458.
336. Kreitman, M. and A. Di Rienzo, *Balancing claims for balancing selection*. Trends Genet, 2004. **20**(7): p. 300-4.
337. Karlsson, E.K., D.P. Kwiatkowski, and P.C. Sabeti, *Natural selection and infectious disease in human populations*. Nat Rev Genet, 2014. **15**(6): p. 379-93.
338. Brunham, R.C., F.A. Plummer, and R.S. Stephens, *Bacterial antigenic variation, host immune response, and pathogen-host coevolution*. Infect Immun, 1993. **61**(6): p. 2273-6.
339. Deitsch, K.W., S.A. Lukehart, and J.R. Stringer, *Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens*. Nat Rev Microbiol, 2009. **7**(7): p. 493-503.
340. Moran, N.A., *Accelerated evolution and Muller's ratchet in endosymbiotic bacteria*. Proc Natl Acad Sci U S A, 1996. **93**(7): p. 2873-8.
341. Fu, Y.X. and W.H. Li, *Statistical tests of neutrality of mutations*. Genetics, 1993. **133**(3): p. 693-709.
342. Conway, D.J., et al., *A principal target of human immunity to malaria identified by molecular population genetic and immunological analyses*. Nat Med, 2000. **6**(6): p. 689-92.
343. Amambua-Ngwa, A., et al., *Population Genomic Scan for Candidate Signatures of Balancing Selection to Guide Antigen Characterization in Malaria Parasites*. PLoS Genetics, 2012. **8**(11): p. e1002992.
344. Yuan, L., et al., *Plasmodium falciparum populations from northeastern Myanmar display high levels of genetic diversity at multiple antigenic loci*. Acta Tropica, 2013. **125**(1): p. 53-59.
345. Mobegi, V.A., et al., *Genome-wide analysis of selection on the malaria parasite Plasmodium falciparum in West African populations of differing infection endemicity*. Mol Biol Evol, 2014. **31**(6): p. 1490-9.
346. Samad, H., et al., *Imputation-based population genetics analysis of Plasmodium falciparum malaria parasites*. PLoS Genet, 2015. **11**(4): p. e1005131.
347. Polley, S.D. and D.J. Conway, *Strong diversifying selection on domains of the Plasmodium falciparum apical membrane antigen 1 gene*. Genetics, 2001. **158**(4): p. 1505-12.
348. Polley, S.D., et al., *Plasmodium falciparum merozoite surface protein 3 is a target of allele-specific immunity and alleles are maintained by natural selection*. J Infect Dis, 2007. **195**(2): p. 279-87.

349. Yahara, K., et al., *Genome-wide survey of codons under diversifying selection in a highly recombining bacterial species, Helicobacter pylori*. DNA Research, 2016. **23**(2): p. 135-143.
350. Thomas, J.C., et al., *Candidate targets of balancing selection in the genome of Staphylococcus aureus*. Molecular Biology and Evolution, 2012. **29**(4): p. 1175-1186.
351. Li, Y., et al., *Distinct effects on diversifying selection by two mechanisms of immunity against Streptococcus pneumoniae*. PLoS Pathogens, 2012. **8**(11): p. e1002989.
352. Hughes, A.L., *Evidence for abundant slightly deleterious polymorphisms in bacterial populations*. Genetics, 2005. **169**(2): p. 533-8.
353. Snoeck, J., et al., *Mapping of positive selection sites in the HIV-1 genome in the context of RNA and protein structural constraints*. Retrovirology, 2011. **8**: p. 87.
354. Li, W., et al., *Positive selection on hemagglutinin and neuraminidase genes of H1N1 influenza viruses*. Virol J, 2011. **8**: p. 183.
355. Joseph, S.J., et al., *Population genomics of Chlamydia trachomatis: insights on drift, selection, recombination, and population structure*. Molecular Biology and Evolution, 2012. **29**(12): p. 3933-3946.
356. Borges, V., et al., *Directional evolution of Chlamydia trachomatis towards niche-specific adaptation*. Journal of Bacteriology, 2012. **194**(22): p. 6143-6153.
357. Chen, C., et al., *In vitro passage selects for Chlamydia muridarum with enhanced infectivity in cultured cells but attenuated pathogenicity in mouse upper genital tract*. Infection and Immunity, 2015.
358. Borges, V., et al., *Chlamydia trachomatis In Vivo to In Vitro Transition Reveals Mechanisms of Phase Variation and Down-Regulation of Virulence Factors*. PLoS One, 2015. **10**(7): p. e0133420.
359. Frikha-Gargouri, O., et al., *Evaluation of an in silico predicted specific and immunogenic antigen from the OmcB protein for the serodiagnosis of Chlamydia trachomatis infections*. BMC Microbiol, 2008. **8**: p. 217.
360. Zhu, S., et al., *Identification of immunodominant linear B-cell epitopes within the major outer membrane protein of Chlamydia trachomatis*. Acta Biochim Biophys Sin (Shanghai), 2010. **42**(11): p. 771-8.
361. Zhu, S., et al., *Identification of linear B-cell epitopes within Tarp of Chlamydia trachomatis*. J Pept Sci, 2014. **20**(12): p. 916-22.
362. Bullock, H.D., S. Hower, and K.A. Fields, *Domain analyses reveal that Chlamydia trachomatis CT694 protein belongs to the membrane-localized family of type III effector proteins*. Journal of Biological Chemistry, 2012. **287**(33): p. 28078-28086.
363. Clifton, D.R., et al., *A chlamydial type III translocated protein is tyrosine-phosphorylated at the site of entry and associated with recruitment of actin*. Proc Natl Acad Sci U S A, 2004. **101**(27): p. 10166-71.
364. Conlan, J.W., I.N. Clarke, and M.E. Ward, *Epitope mapping with solid-phase peptides: identification of type-, subspecies-, species- and genus-reactive antibody binding domains on the major outer membrane protein of Chlamydia trachomatis*. Molecular Microbiology, 1988. **2**(5): p. 673-679.
365. Su, H., G.J. Spangrude, and H.D. Caldwell, *Expression of Fc gamma RIII on HeLa 229 cells: possible effect on in vitro neutralization of Chlamydia trachomatis*. Infection and Immunity, 1991. **59**(10): p. 3811-3814.
366. Peterson, E.M., et al., *Effect of immunoglobulin G isotype on the infectivity of Chlamydia trachomatis in a mouse model of intravaginal infection*. Infect Immun, 1997. **65**(7): p. 2693-9.

367. Tetteh, K.K., et al., *Prospective identification of malaria parasite genes under balancing selection*. PLoS One, 2009. **4**(5): p. e5568.
368. Wang, I.N., et al., *Genetic diversity of ospC in a local population of Borrelia burgdorferi sensu stricto*. Genetics, 1999. **151**(1): p. 15-30.
369. Lutter, E.I., C. Martens, and T. Hackstadt, *Evolution and conservation of predicted inclusion membrane proteins in chlamydiae*. Comp Funct Genomics, 2012. **2012**: p. 362104.
370. Li, Z., et al., *Characterization of fifty putative inclusion membrane proteins encoded in the Chlamydia trachomatis genome*. Infection and Immunity, 2008. **76**(6): p. 2746-2757.
371. Newhall, W.J.t., *Biosynthesis and disulfide cross-linking of outer membrane components during the growth cycle of Chlamydia trachomatis*. Infect Immun, 1987. **55**(1): p. 162-8.
372. Zhang, Y.X., et al., *The low-molecular-mass, cysteine-rich outer membrane protein of Chlamydia trachomatis possesses both biovar- and species-specific epitopes*. Infection and Immunity, 1987. **55**(11): p. 2570-2573.
373. Clarke, I.N., M.E. Ward, and P.R. Lambden, *Molecular cloning and sequence analysis of a developmentally regulated cysteine-rich outer membrane protein from Chlamydia trachomatis*. Gene, 1988. **71**(2): p. 307-14.
374. de la Maza, L.M., et al., *Sequence diversity of the 60-kilodalton protein and of a putative 15-kilodalton protein between the trachoma and lymphogranuloma venereum biovars of Chlamydia trachomatis*. Infect Immun, 1991. **59**(3): p. 1196-201.
375. Borges, V., *Evolutionary dynamics of Chlamydia trachomatis genome and identification of molecular patterns of hypothetical protein coding genes*, in Faculdade de Ciências e Tecnologia. 2015, Universidade Nova de Lisboa.
376. Weber, M.M., et al., *Expression and localization of predicted inclusion membrane proteins in Chlamydia trachomatis*. Infection and Immunity, 2015.
377. Griffiths, E., M.S. Ventresca, and R.S. Gupta, *BLAST screening of chlamydial genomes to identify signature proteins that are unique for the Chlamydiales, Chlamydiaceae, Chlamydophila and Chlamydia groups of species*. BMC Genomics, 2006. **7**: p. 14.
378. Loomis, W.P. and M.N. Starnbach, *Chlamydia trachomatis infection alters the development of memory CD8+ T cells*. J Immunol, 2006. **177**(6): p. 4021-7.
379. Borges, V., et al., *Complete Genome Sequence of Chlamydia trachomatis Ocular Serovar C Strain TW-3*. Genome Announcements, 2014. **2**(1): p. e01204-13.
380. Hayes, L.J., et al., *Genotyping of Chlamydia trachomatis from a trachoma-endemic village in the Gambia by a nested polymerase chain reaction: identification of strain variants*. J Infect Dis, 1992. **166**(5): p. 1173-7.
381. Takourt, B., et al., *Direct genotyping and nucleotide sequence analysis of VS1 and VS2 of the Omp1 gene of Chlamydia trachomatis from Moroccan trachomatous specimens*. Microbes Infect, 2001. **3**(6): p. 459-66.
382. Miyairi, I., et al., *Different growth rates of Chlamydia trachomatis biovars reflect pathotype*. J Infect Dis, 2006. **194**(3): p. 350-7.
383. Borges, V., et al., *Normalization strategies for real-time expression data in Chlamydia trachomatis*. J Microbiol Methods, 2010. **82**(3): p. 256-64.
384. Jorgensen, I. and R.H. Valdivia, *Pmp-like proteins Pls1 and Pls2 are secreted into the lumen of the Chlamydia trachomatis inclusion*. Infect Immun, 2008. **76**(9): p. 3940-50.
385. Mu, F.-T., et al., *EEA1, an Early Endosome-Associated Protein*. The Journal of Biological Chemistry, 1995. **270**(22): p. 13503-13511.

386. McClellan, D.A. and K.G. McCracken, *Estimating the influence of selection on the variable amino acid sites of the cytochrome B protein functional domains*. Mol Biol Evol, 2001. **18**(6): p. 917-25.
387. Spielman, S.J. and C.O. Wilke, *Membrane environment imposes unique selection pressures on transmembrane domains of G protein-coupled receptors*. J Mol Evol, 2013. **76**(3): p. 172-82.
388. Shaw, E.I., et al., *Three temporal classes of gene expression during the Chlamydia trachomatis developmental cycle*. Mol Microbiol, 2000. **37**(4): p. 913-25.
389. Simonsen, A., et al., *EEA1 links PI(3)K function to Rab5 regulation of endosome fusion*. Nature, 1998. **394**(6692): p. 494-498.
390. Pfeffer, S.R., *Rab GTPases: specifying and deciphering organelle identity and function*. Trends Cell Biol, 2001. **11**(12): p. 487-91.
391. Tuvim, M.J., et al., *Traffic control: Rab GTPases and the regulation of interorganellar transport*. News Physiol Sci, 2001. **16**: p. 56-61.
392. Fields, K.A. and T. Hackstadt, *The chlamydial inclusion: escape from the endocytic pathway*. Annual Review of Cell and Developmental Biology, 2002. **18**: p. 221-245.
393. Rzomp, K.A., A.R. Moorhead, and M.A. Scidmore, *The GTPase Rab4 interacts with Chlamydia trachomatis inclusion membrane protein CT229*. Infection and Immunity, 2006. **74**(9): p. 5362-5373.
394. Capmany, A., N. Leiva, and M.T. Damiani, *Golgi-associated Rab14, a new regulator for Chlamydia trachomatis infection outcome*. Commun Integr Biol, 2011. **4**(5): p. 590-3.
395. Bucci, C., et al., *Rab7: a key to lysosome biogenesis*. Mol Biol Cell, 2000. **11**(2): p. 467-80.
396. Priya, A., et al., *Molecular insights into Rab7-mediated endosomal recruitment of core retromer: deciphering the role of Vps26 and Vps35*. Traffic, 2015. **16**(1): p. 68-84.
397. Zhang, M., et al., *Rab7: roles in membrane trafficking and disease*. Biosci Rep, 2009. **29**(3): p. 193-209.
398. Sun, H.S., et al., *Chlamydia trachomatis vacuole maturation in infected macrophages*. J Leukoc Biol, 2012. **92**(4): p. 815-27.
399. van Ooij, C., G. Apodaca, and J. Engel, *Characterization of the Chlamydia trachomatis vacuole and its interaction with the host endocytic pathway in HeLa cells*. Infect Immun, 1997. **65**(2): p. 758-66.
400. Patel, A.L., et al., *Activation of epidermal growth factor receptor is required for Chlamydia trachomatis development*. BMC Microbiol, 2014. **14**: p. 277.
401. Yoon, H.Y., J.S. Lee, and P.A. Randazzo, *ARAP1 regulates endocytosis of EGFR*. Traffic, 2008. **9**(12): p. 2236-52.
402. Ceresa, B.P. and S.J. Bahr, *rab7 activity affects epidermal growth factor:epidermal growth factor receptor degradation by regulating endocytic trafficking from the late endosome*. J Biol Chem, 2006. **281**(2): p. 1099-106.
403. Roberts, C., et al., *Conjunctival fibrosis and the innate barriers to Chlamydia trachomatis intracellular infection: a genome wide association study*. Sci Rep, 2015. **5**: p. 17447.
404. Thalmann, J., et al., *Actin re-organization induced by Chlamydia trachomatis serovar D-evidence for a critical role of the effector protein CT166 targeting Rac*. PLoS One, 2010. **5**(3): p. e9887.
405. Frikha-Gargouri, O., et al., *Diagnostic value of enzyme-linked immunosorbent assays using hypothetical proteins CT226 and CT795 as antigens in Chlamydia*

- trachomatis* serodiagnosis. Diagnostic Microbiology and Infectious Disease, 2009. **65**(3): p. 224-231.
406. Mueller, K.E., G.V. Plano, and K.A. Fields, *New frontiers in type III secretion biology: the Chlamydia perspective*. Infect Immun, 2014. **82**(1): p. 2-9.
 407. Misaghi, S., et al., *Chlamydia trachomatis*-derived deubiquitinating enzymes in mammalian cells during infection. Mol Microbiol, 2006. **61**(1): p. 142-50.
 408. Le Negrate, G., et al., *ChlaDub1 of Chlamydia trachomatis suppresses NF-kappaB activation and inhibits IkappaBalpha ubiquitination and degradation*. Cell Microbiol, 2008. **10**(9): p. 1879-92.
 409. Nunes, A., et al., *Bioinformatic Analysis of Chlamydia trachomatis Polymorphic Membrane Proteins PmpE, PmpF, PmpG and PmpH as Potential Vaccine Antigens*. PLoS One, 2015. **10**(7): p. e0131695.
 410. Dieterle, S. and J. Wollenhaupt, *Humoral immune response to the chlamydial heat shock proteins hsp60 and hsp70 in Chlamydia-associated chronic salpingitis with tubal occlusion*. Hum Reprod, 1996. **11**(6): p. 1352-6.
 411. Raulston, J.E., et al., *Surface accessibility of the 70-kilodalton Chlamydia trachomatis heat shock protein following reduction of outer membrane protein disulfide bonds*. Infect Immun, 2002. **70**(2): p. 535-43.
 412. Badamchi-Zadeh, A., et al., *Intramuscular Immunisation with Chlamydial Proteins Induces Chlamydia trachomatis Specific Ocular Antibodies*. PLoS One, 2015. **10**(10): p. e0141209.
 413. Walraven, G., et al., *The burden of reproductive-organ disease in rural women in The Gambia, West Africa*. Lancet, 2001. **357**(9263): p. 1161-7.
 414. Gueye Ndiaye, A., et al., *[Screening for HIV, syphilis, Chlamydia trachomatis and Neisseria gonorrhoeae during a combined survey conducted in Malicouna, a Senegalese rural area]*. Bull Soc Pathol Exot, 2009. **102**(3): p. 150-4.
 415. Newman, L., et al., *Global Estimates of the Prevalence and Incidence of Four Curable Sexually Transmitted Infections in 2012 Based on Systematic Review and Global Reporting*. PLoS One, 2015. **10**(12): p. e0143304.
 416. Tuuminen, T., P. Palomaki, and J. Paavonen, *The use of serologic tests for the diagnosis of chlamydial infections*. Journal of Microbiological Methods, 2000. **42**(3): p. 265-279.
 417. Cohen, C.R., et al., *Immunoepidemiologic Profile of Chlamydia trachomatis Infection: Importance of Heat-Shock Protein 60 and Interferon-gamma*. Journal of Infectious Diseases, 2005. **192**(4): p. 591-599.
 418. Russell, A.N., et al., *Identification of Chlamydia trachomatis antigens recognized by T cells from highly exposed women who limit or resist genital tract infection*. Journal of Infectious Diseases, 2016.
 419. Taylor, H.R., et al., *Longitudinal study of the microbiology of endemic trachoma*. J Clin Microbiol, 1991. **29**(8): p. 1593-5.
 420. de la Maza, L.M., et al., *Intravaginal inoculation of mice with the Chlamydia trachomatis mouse pneumonitis biovar results in infertility*. Infection and Immunity, 1994. **62**(5): p. 2094-2097.
 421. Darville, T., et al., *Mouse strain-dependent variation in the course and outcome of chlamydial genital tract infection is associated with differences in host response*. Infection and Immunity, 1997. **65**(8): p. 3065-3073.
 422. Lyons, J.M., et al., *Comparison of multiple genital tract infections with Chlamydia trachomatis in different strains of female mice*. J Microbiol Immunol Infect, 2005. **38**(6): p. 383-93.
 423. Teng, A., et al., *Proteomic identification of immunodominant chlamydial antigens in a mouse model*. Journal of Proteomics, 2012. **77**: p. 176-186.

424. Holland, M.J., et al., *The frequency of Chlamydia trachomatis major outer membrane protein-specific CD8+ T lymphocytes in active trachoma is associated with current ocular infection*. Infection and Immunity, 2006. **74**(3): p. 1565-1572.

Appendix

Buffers

Blocking buffer A; 0.05 % Triton X-100 and 2.5 % skimmed milk in phosphate-buffered saline (PBS) at pH 7.5.

Blocking buffer B; 0.05 % Triton X-100 and ten % skimmed milk in PBS at pH 7.5.

Blocking buffer C; ten % *Escherichia coli* lysate (McLab, San Francisco, CA) in protein array blocking buffer (Whatman, Piscataway, NJ).

Blocking buffer D; 0.05 % Triton X-100 in Tris-buffered saline (TBS) at pH 7.4.

Cleavage buffer; 50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA and 1 mM DTT in PBS at pH 7.5.

Co-immunoprecipitation buffer; 25 mM Tris-HCl, 1 mM EDTA, 1 mM EGTA, 100 mM NaCl, 0.1 % Triton X-100, 0.1 % NP40, 0.1 % ASB-14 and 1 % Protease inhibitor cocktail (Sigma Aldrich) in H₂O.

Coating buffer A; 0.05 M carbonate-bicarbonate in PBS at pH 9.6.

Coating buffer B; 0.1 % bovine serum albumin (BSA) and 0.1 % sodium azide in PBS at pH 7.5.

Culture medium A; 10 % foetal calf serum (FCS) and 1 % PenStrep in DMEM.

Elution buffer A; 50 mM Tris-HCl and 10 mM reduced glutathione in PBS at pH 8.

Elution buffer B; 100 mM Tris-HCl, 200 mM EDTA and 2 % sodium dodecyl sulphate (SDS) in H₂O.

Infection medium; 10 % FCS and 1 % gentamycin in DMEM.

Laemmli buffer; four % SDS, 20 % glycerol, 120 mM Tris-Cl (pH 6.8) and 0.02 % bromophenol blue in H₂O

Transfection medium; 1.4 % TurboFect (Thermo Fisher Scientific) in DMEM.

Triton lysis buffer (Sharma 2006); 1 % Triton X-100, 1 mM phenylmethylsulfonyl fluoride (PMSF), 75 IU/ml aprotinin, 20 µM leupeptin and 1.6 µM pepstatin in PBS at pH 7.5.

Washing buffer A; 0.05 % Triton X-100 in PBS at pH 7.5.

Washing buffer B; 0.05 % Triton X-100 and 5 % skimmed milk in TBS at pH 7.4.

Table 0.1. Summary of previous Ct micro-array antigen identification.

Ct D/UW3 nomenclature was used. T and B-cell recognition numbers were the how often they were found to be immunogenic in 8 previous Ct micro-arrays.

ID	NAME	T-CELL RECOGNITION	B-CELL RECOGNITION
CT004	GatB	1	0
CT015	PhoH	1	0
CT016	Hypothetical	1	0
CT019	IleS	0	1
CT022	RpmE2	0	1
CT035	BPL	1	0
CT043	Hypothetical	2	0
CT049	Hypothetical	0	1
CT067	YtgA	0	2
CT082	Hypothetical	0	1
CT089	CopN	0	6
CT101	Hypothetical	0	1
CT110	GroEL1 (HSP60)	1	4
CT111	GroES	1	0
CT114	Hypothetical	1	0
CT116	IncE	0	1
CT117	IncF	0	1

CT118	IncG	0	1
CT119	IncA	1	4
CT142	Hypothetical	0	2
CT143	Hypothetical	0	2
CT147	Hypothetical	0	4
CT153	Hypothetical	1	2
CT168	Hypothetical	1	0
CT184	YqgF	1	0
CT226	Hypothetical	0	1
CT228	Hypothetical	0	1
CT240	RecR	0	1
CT255	Hypothetical	1	0
CT279	Nqr3	1	0
CT301	PknD	0	1
CT315	RpoB	1	0
CT316	R17	0	1
CT322	TuEF	1	3
CT341	DnaJ	1	0
CT342	Rs21	1	0
CT355	Hypothetical	0	1
CT372	Hypothetical	1	0
CT376	MdhC	0	2
CT381	ArtJ	0	4
CT396	DnaK (HSP70)	2	1
CT414	PmpC	0	2

CT415	YebL	0	1
CT442	CrpA	0	4
CT443	OmcB	3	6
CT456	TARP	0	3
CT460	SWIB	1	0
CT480	DppA	1	0
CT492	YacE	1	0
CT509	RS13	1	0
CT529	CapI	0	3
CT553	Fmu	0	1
CT556	Hypothetical	0	1
CT557	LpdA	0	2
CT559	YscJ	0	1
CT571	GspE	0	1
CT587	Eno	1	0
CT589	Hypothetical	0	1
CT600	Pal	1	0
CT601	PapQ	1	0
CT603	TSA	1	1
CT611	Hypothetical	1	0
CT619	Hypothetical	0	1
CT622	CHLPN homologue	0	1
CT667	Hypothetical	0	1
CT681	MOMP	2	5
CT694	Hypothetical	1	3

CT695	Hypothetical	0	3
CT702	Hypothetical	0	1
CT706	ClpP2	0	1
CT709	MreB	0	1
CT711	Hypothetical	1	0
CT716	Hypothetical	1	0
CT733	Hypothetical	1	0
CT734	Hypothetical	1	0
CT755	GroEL3	1	1
CT795	Hypothetical	0	3
CT798	GlgA	0	2
CT806	Ptr	0	2
CT812	PmpD	1	2
CT813	Hypothetical	0	2
CT823	HtrA	1	2
CT828	NrdB	0	1
CT841	FtsH	0	1
CT858	CPAF	1	4
CT866	GlgB	0	1
CT871	PmpG	1	0
CT872	PmpH	1	0
CT875	Hypothetical	1	3
PCT03	Pgp3	0	2

Table 0.2. CT442-his primers for cloning into pET22(b)+.

RESTRICTION SEQUENCE (RESTRICTION SITE UNDERLINED)
SITE

BAMHI	GGCGCC <u>GGATCC</u> ATGAGCACTGTACCCGTTGTTCAAG
XHOI	GGCCGC <u>CTCGAG</u> TTGGGTCTGATCCACCAGACTATTTC

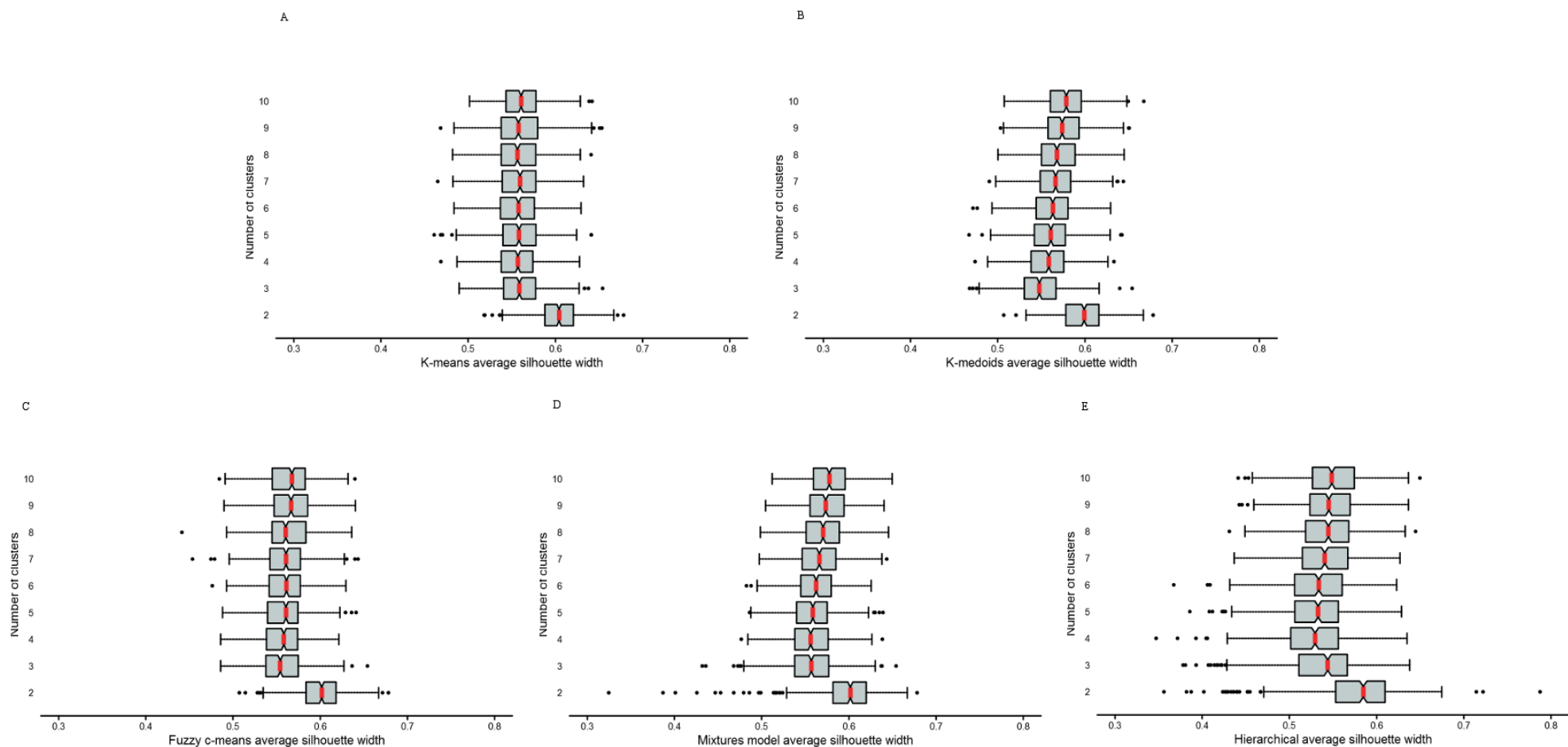


Figure 1. Average silhouette width comparison of number of clusters for chapter 4.

Number of clusters are on the left-hand side, from 2 to 10. Clustering methods trialled were A) K-means, B) K-medoids, C) Fuzzy c-means, D) Mixtures model and E) Hierarchical clustering.

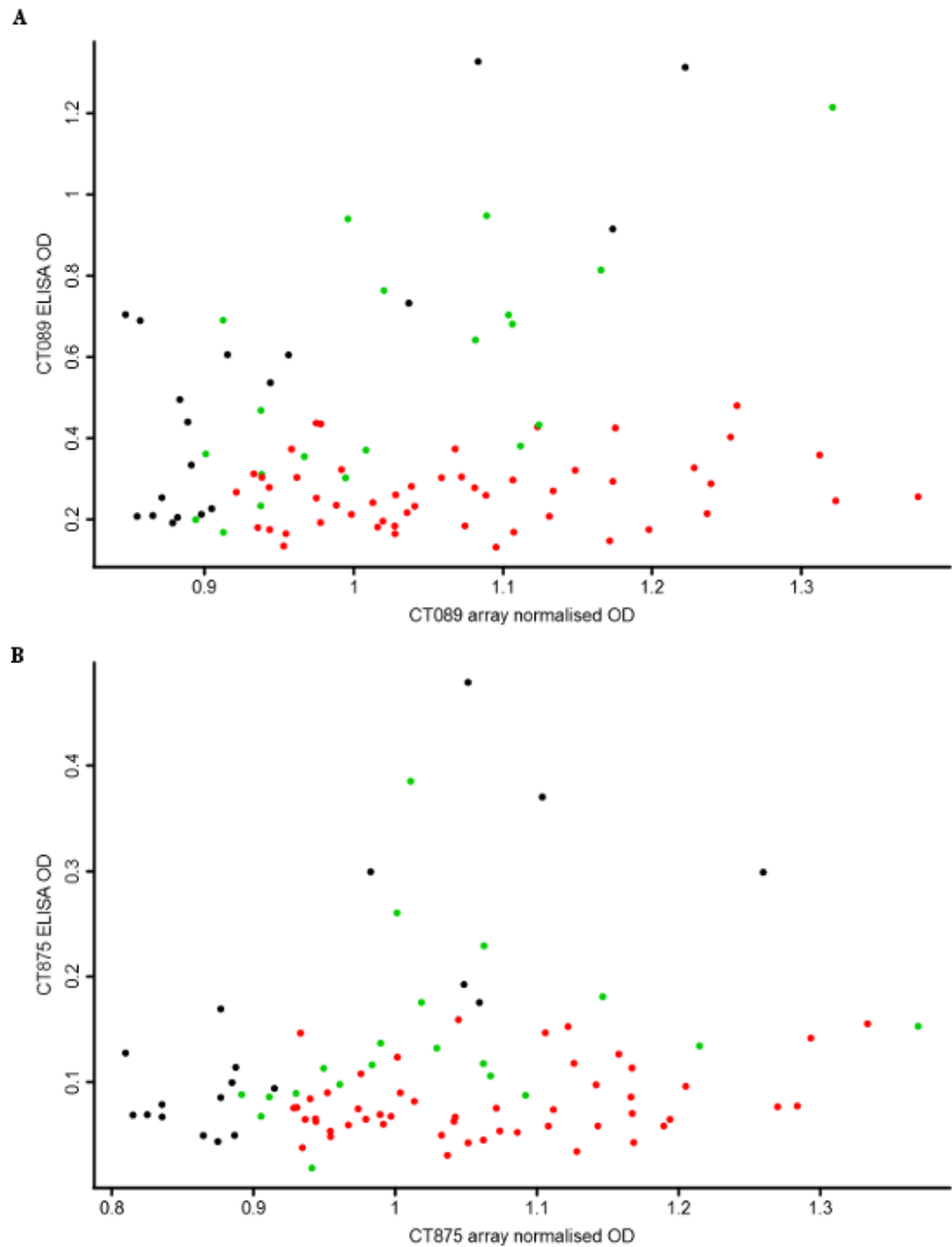


Figure 2. Comparison of ELISA and array results for 90 longitudinal sera.

Results were compared and sera with discordant results for CT089 and CT875 respectively were indicated (both discordant, red), or discordant for just one of CT089 or CT875 (one discordant, green).

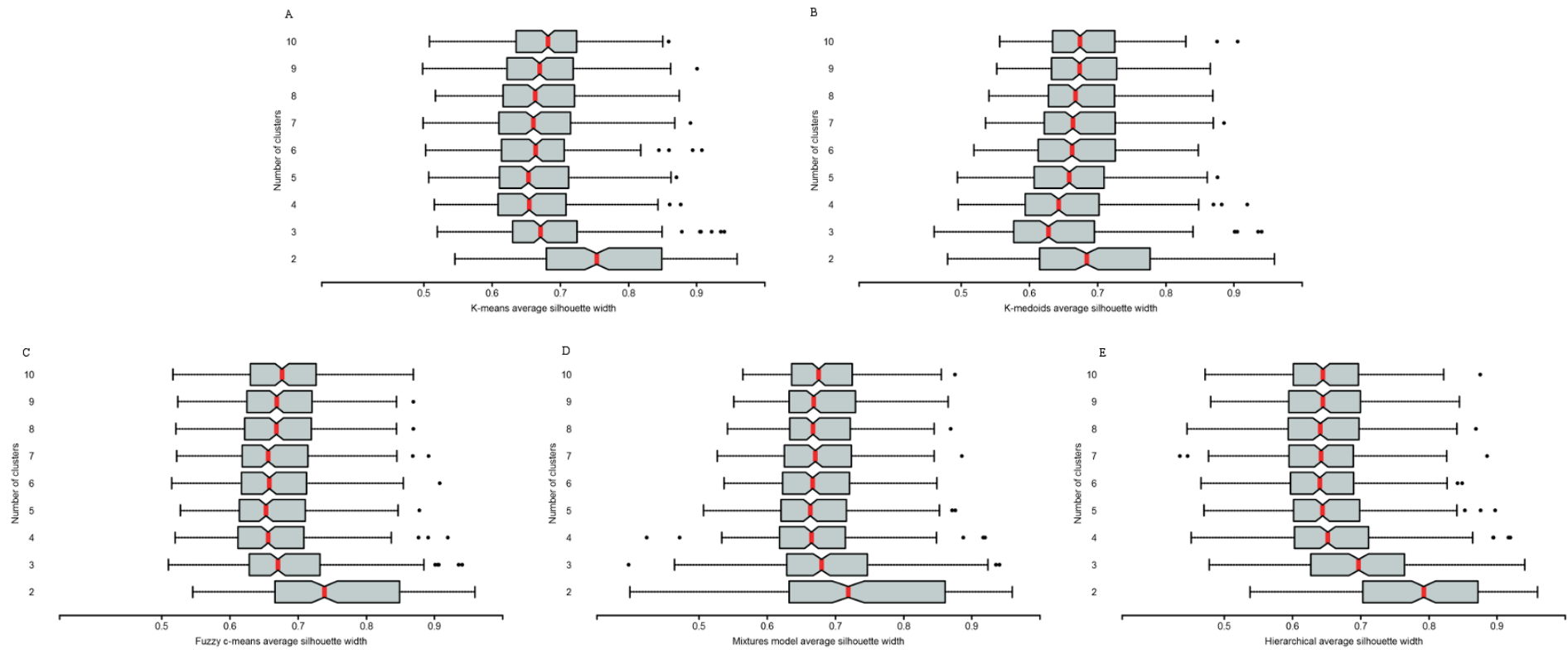


Figure 3. Average silhouette width comparison of number of clusters for chapter 5.

Number of clusters are on the left-hand side, from 2 to 10. Clustering methods trialled were A) K-means, B) K-medoids, C) Fuzzy c-means, D) Mixtures model and E) Hierarchical clustering.

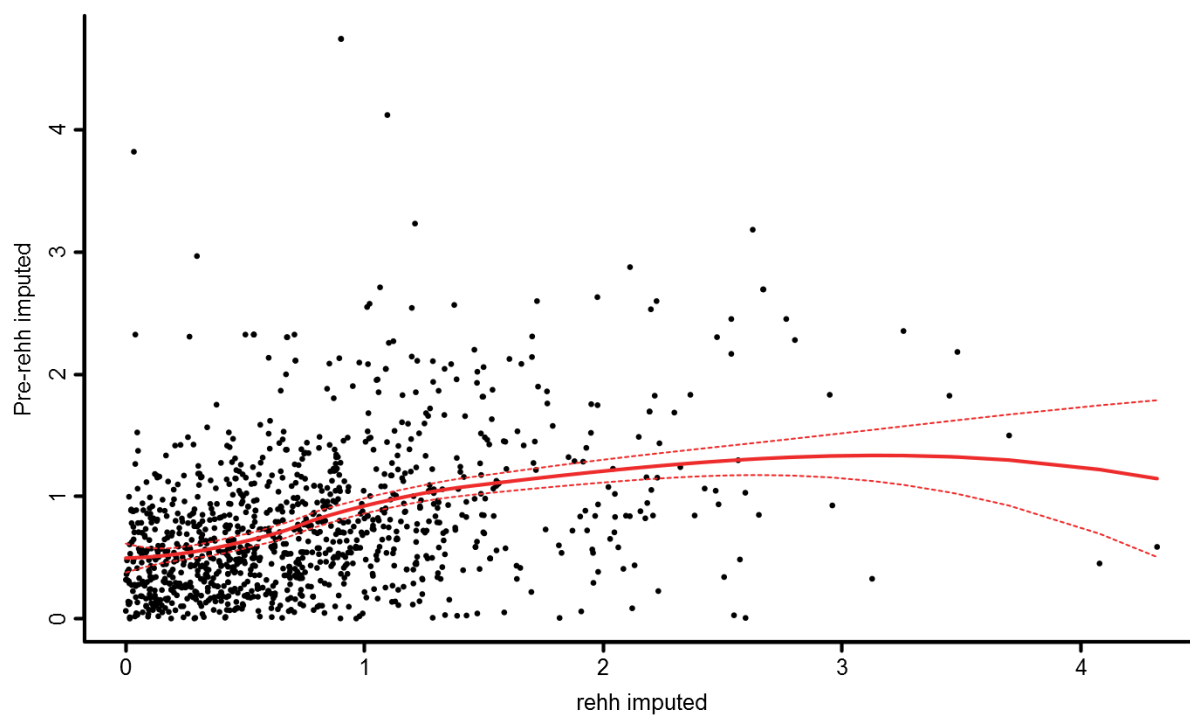


Figure 4. Integrated haplotype scores pre- and post-imputation.

A loess regression was fitted to compare the data (red line), with 95 % confidence intervals (dashed red lines).

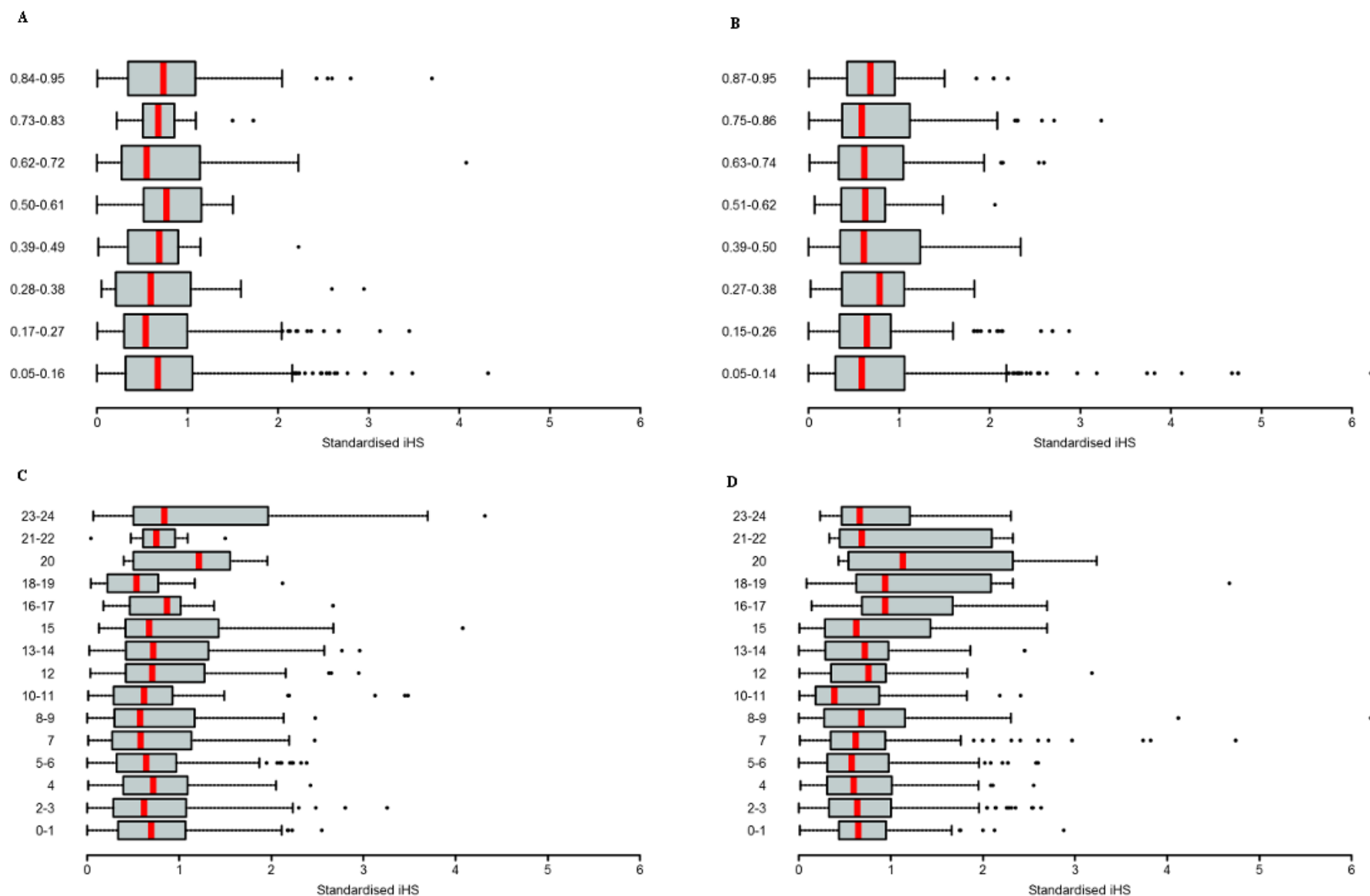


Figure 5. Comparison of iHS scores by minor allele frequency (MAF) and percentage missed calls.

The distribution of iHS scores by MAF (A and B) and percentage missed calls (C and D) were compared pre- (A and C) and post-imputation (B and D) of missed calls. There was no systematic difference in scores by MAF. Scores increased above 15 % missed calls post-imputation, these SNPs were excluded.

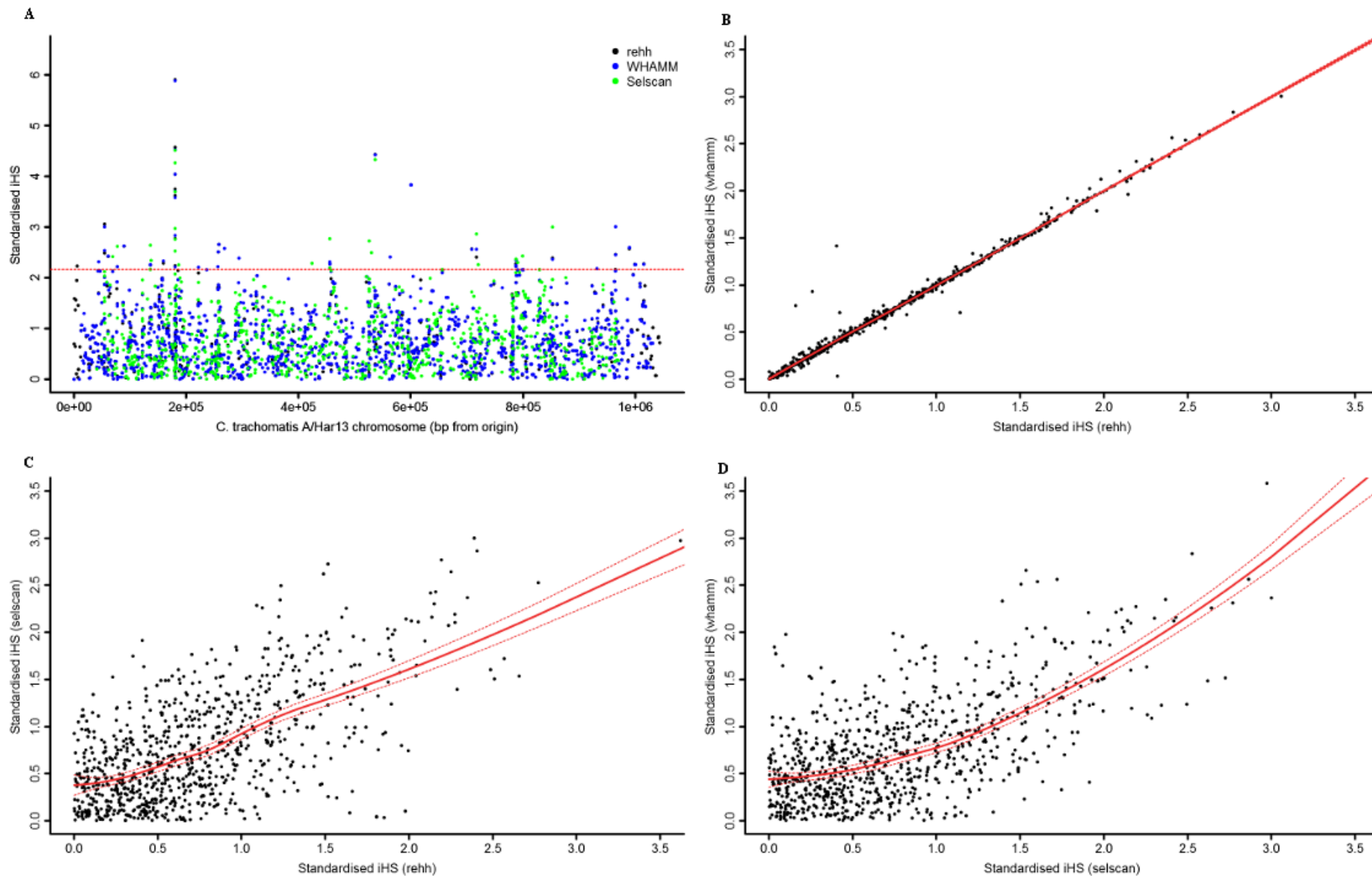


Figure 6. Comparison of iHS scores from different software.

High scoring regions were common using all software (A); rehh (black), WHAMM (blue) and Selscan (green). Correlation between rehh and WHAMM was strong (B). Correlation between Selscan and rehh (C) and WHAMM (D) was moderate. A loess regression was fitted to compare the data (red line), with 95 % confidence intervals (dashed red lines).

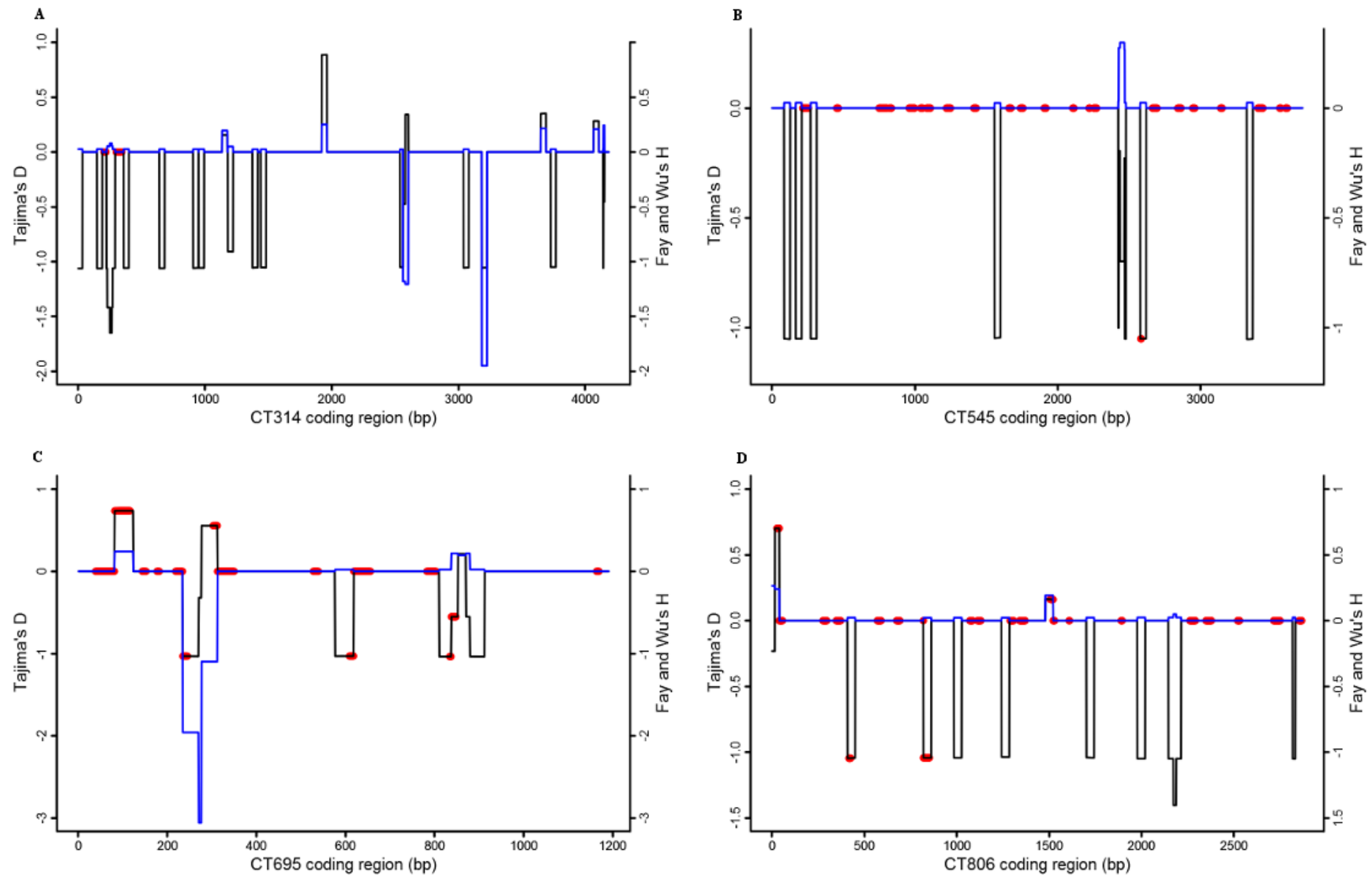


Figure 7. Evidence of selection in CT314, CT545, CT695, CT806.

Values of D (black) and H (blue) across the gene based on sliding windows of 42 nucleotides are shown. Predicted epitopes are indicated (red). A) CT314. B) CT545. C) CT695. D) CT806.

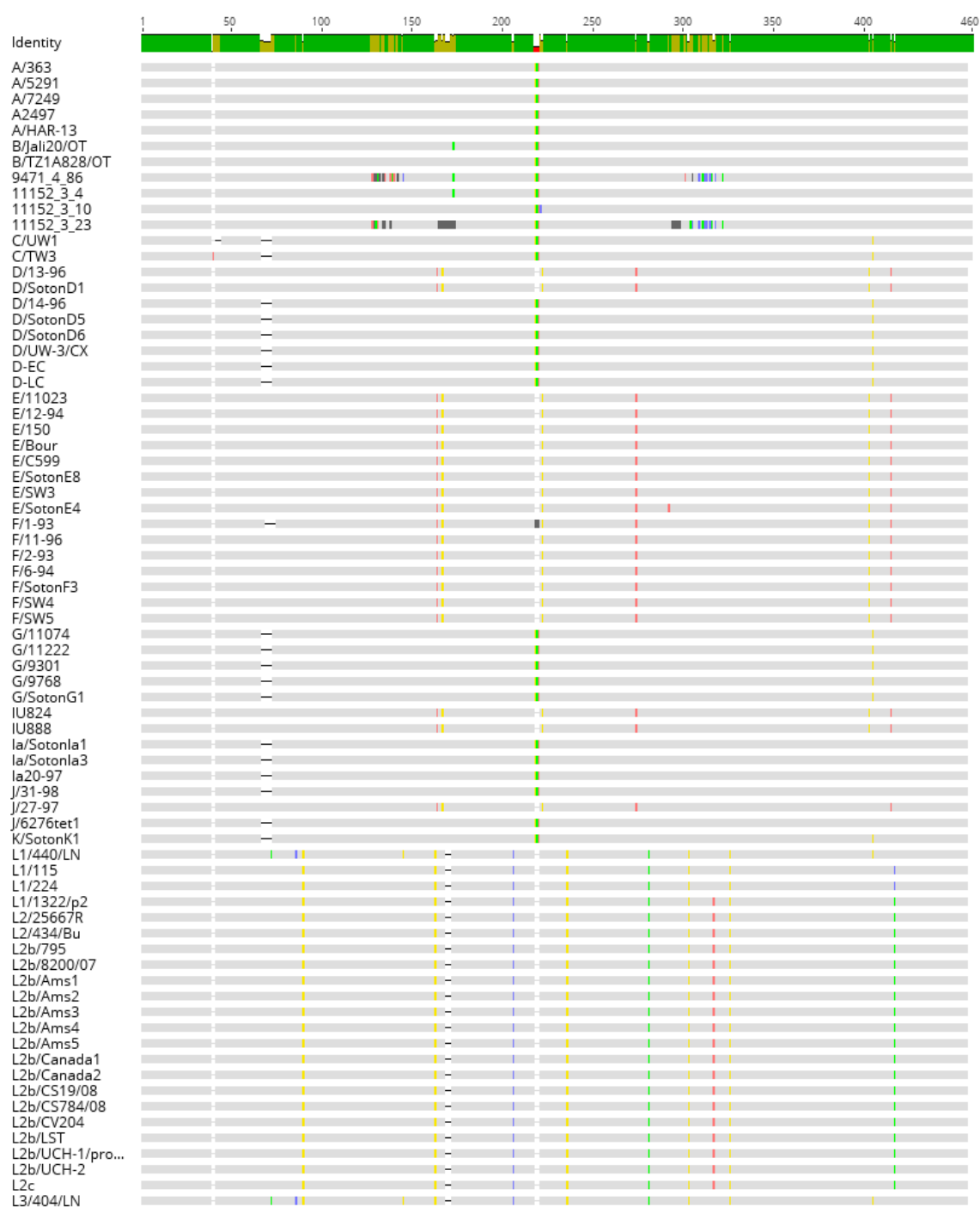


Figure 8. Nucleotide alignment of CT442 from serovar-representative strains.

The ‘identity’ sequence shows relative conservation of sequence, from identical (green) to increasing levels of variation (lighter shades of green). This was produced using Geneious.

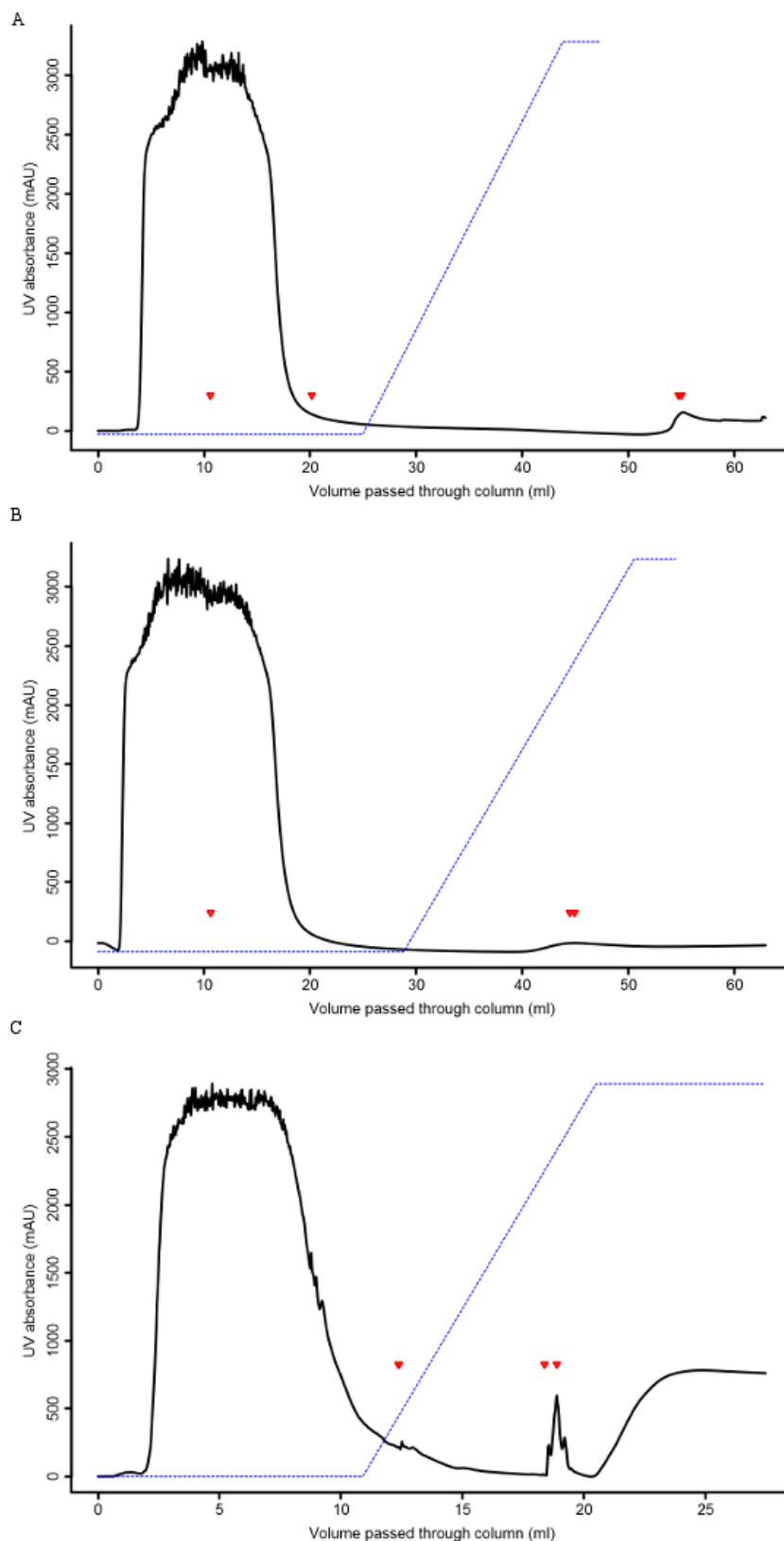


Figure 9. ÄKTA size exclusion-based chromatograms for CT442-His.

UV absorbance (y-axis) shows amount of protein being eluted, concentration of imidazole was indicated (blue line, 0-100 %). A) One wash and two peaks (EL1 and EL2 from Figure 7.6). B) Two washes and two peaks (EL3 and EL4 from Figure 7.6). C) Two peaks (Figure 7.7).

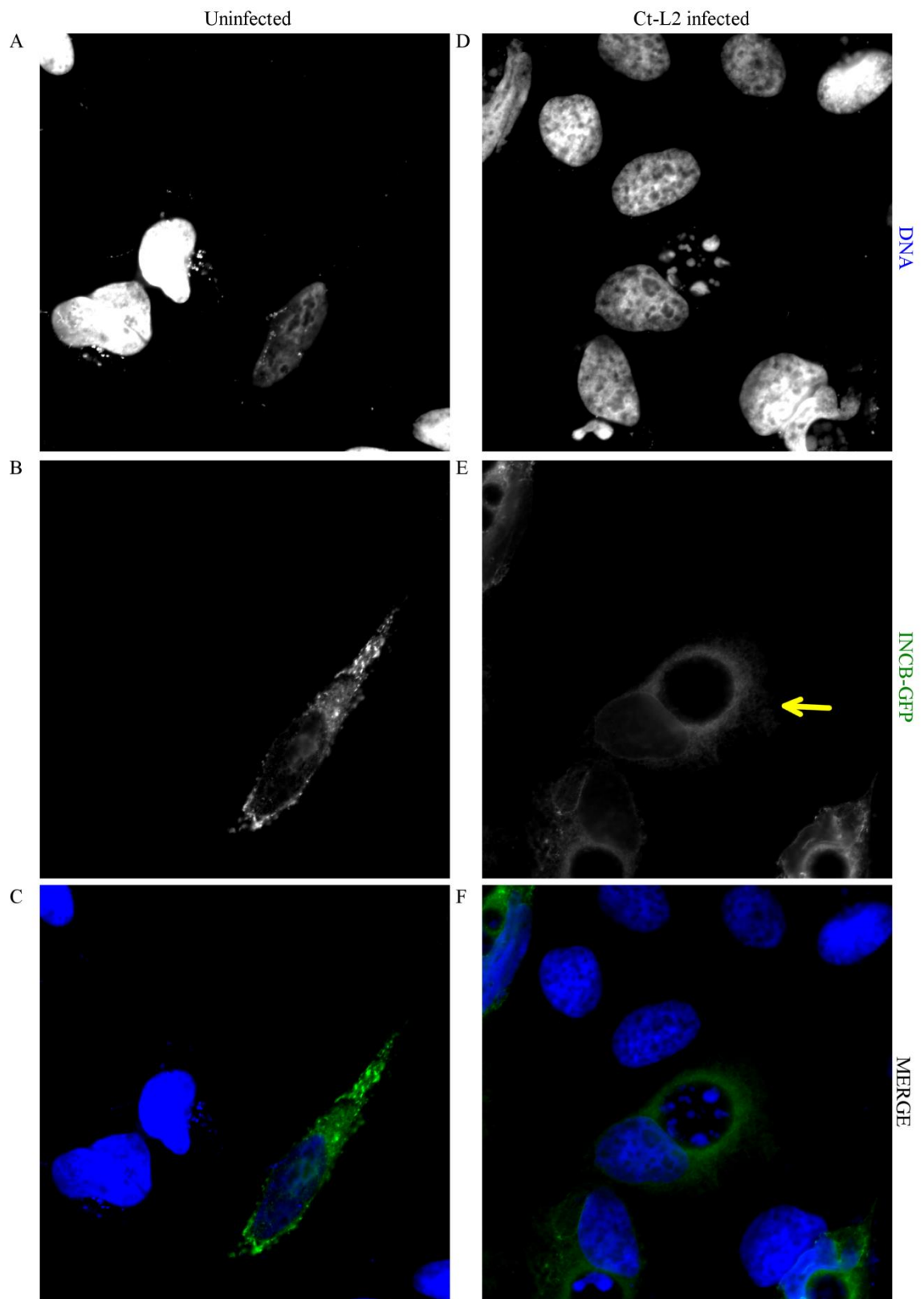


Figure 10. Localisation of IncB-GFP in HeLa cells.

HeLa cells were transfected with IncB-GFP and fixed after 24 hours or then infected with Ct-L2 and fixed 24 HPI. Cells were stained for DNA (A, blue in merged panel) and GFP (B, green in merged panel). Inclusions are indicated (yellow arrows).

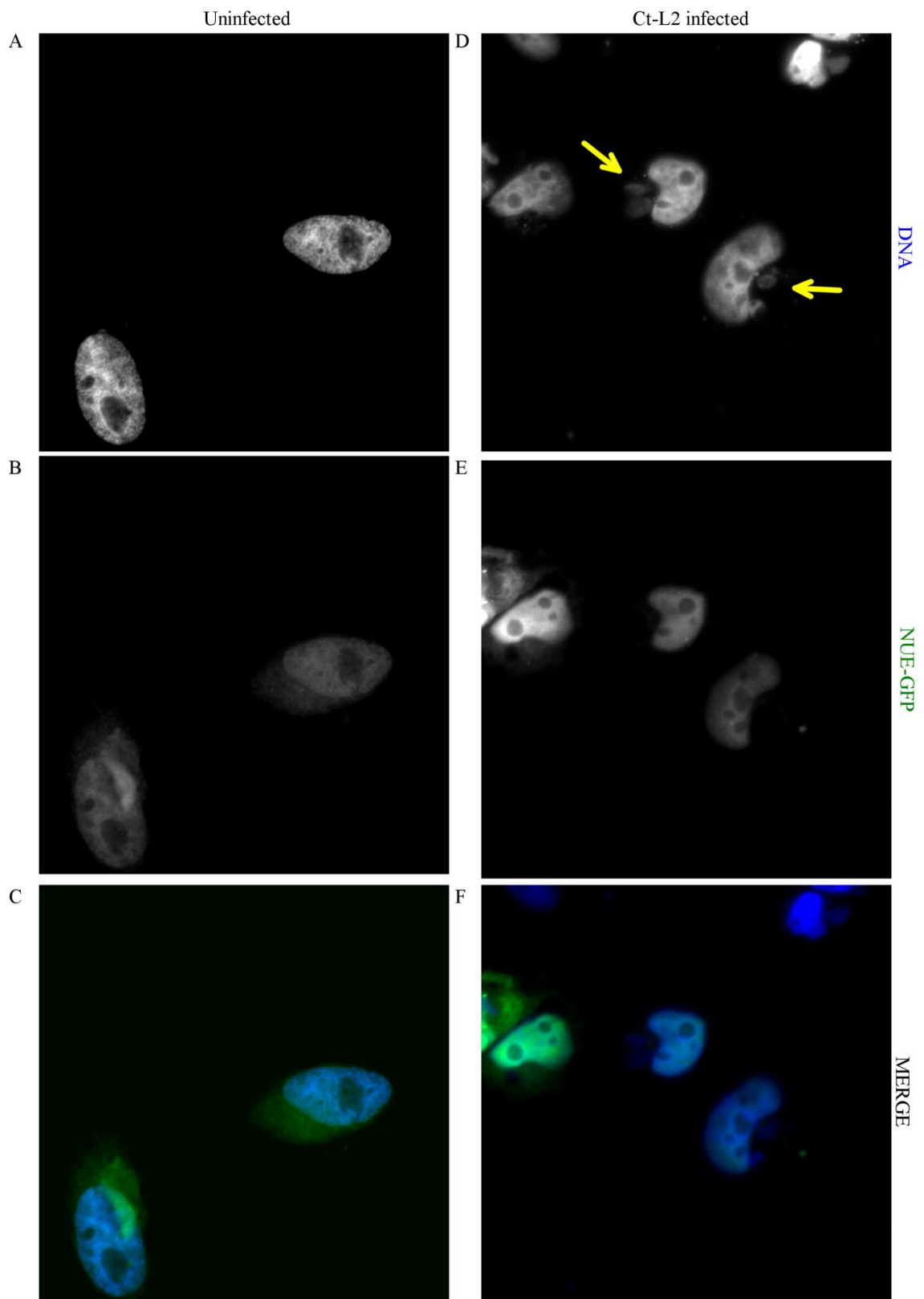


Figure 11. Localisation of NUE-GFP in HeLa cells.

HeLa cells were transfected with NUE-GFP and fixed after 24 hours or then infected with Ct-L2 and fixed 24 HPI. Cells were stained for DNA (A, blue in merged panel) and GFP (B, green in merged panel). Inclusions are indicated (yellow arrows).